Generation and Analysis of Verbal Route Directions for Blind Navigation

John Nicholson

Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Computer Sciences Commons

Recommended Citation
Nicholson, John, "Generation and Analysis of Verbal Route Directions for Blind Navigation" (2010). All Graduate Theses and Dissertations. 672.
https://digitalcommons.usu.edu/etd/672
GENERATION AND ANALYSIS OF VERBAL ROUTE
DIRECTIONS FOR BLIND NAVIGATION

by

John Nicholson

A dissertation submitted in partial fulfillment
of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
Computer Science

Approved:

Dr. Vladimir A. Kulyukin
Major Professor

Dr. Donald H. Cooley
Committee Member

Dr. Daniel C. Coster
Committee Member

Dr. Nicholas Flann
Committee Member

Dr. Minghui Jiang
Committee Member

Dr. Byron R. Burnham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah
2010
According to the National Federation of the Blind, there are an estimated 10 million people in the United States who are visually impaired. Of these, 1.3 million are legally blind. Many people with extreme vision loss receive orientation and mobility training in order to help them learn skills that allow them to travel and navigate multiple types of indoor and outdoor environments. Even with this training, a fundamental problem these people face is learning new routes, especially in environments with which they are not familiar. Although the research community has developed a number of localization and navigation aids that are meant to provide navigation assistance, only a handful have reached the marketplace, and the adoption rate for these devices remains low.

Most assistive navigation devices take responsibility for the navigation and localization processes, leaving the user only to respond to the devices’ commands. This thesis takes a different approach and proposes that because of the high level of navigation ability achieved through years of training and everyday travel, the navigation skills of people with visual impairments should be considered an integral part of the navigation system. People with visual impairments are capable of following natural language instructions similar to those given by a visually impaired person communicating route directions over the phone.
to another person with visual impairments. Devices based on this premise can be built, delivering only verbal route descriptions. As a result, it is not necessary to install complex sensors in the environment.

This thesis has four hypotheses that are addressed by two systems. The first hypothesis is that a navigational assistance system for the blind can leverage the skills and abilities of the visually impaired, and does not necessarily need complex sensors embedded in the environment to succeed. The second hypothesis is that verbal route descriptions are adequate for guiding a person with visual impairments when shopping in a supermarket for products located in aisles on shelves. These two hypotheses are addressed by ShopTalk, a system which helps blind users shop independently in a grocery store using verbal route descriptions.

The third hypothesis is that information extraction techniques can be used to extract landmarks from natural language route descriptions. The fourth and final hypothesis is that new natural language route descriptions can be inferred from a set of landmarks and a set of natural language route descriptions whose statements have been tagged with landmarks from the landmark set. These two hypotheses are addressed by the Route Analysis Engine, an information extraction-based system for analyzing natural language route descriptions.
ACKNOWLEDGMENTS

I would first like to thank my adviser, Dr. Vladimir Kulyukin, for guiding me through the PhD process. It is largely because of him that I completed this work. I thank my committee members, Dr. Daniel Coster, Dr. Donald Cooley, Dr. Nick Flann, and Dr. Minghui Jiang, for their helpful feedback. I would like to offer special thanks to Dr. Coster for helping with the statistical analysis of the experimental data. I also thank my fellow PhD student Aliasgar Kutiyanawala and the other occupants of Old Main Room 405 for listening to my frequent rants and raves. I would like to thank Mr. Sachin Pavithran, a visually impaired training and development specialist at the USU Center for Persons with Disabilities, for his valuable feedback on the devices and experiments and insight into the issues faced by people with visual impairments. Many thanks to the multiple participants who graciously helped to test our various pieces of technology. I also gratefully acknowledge and thank Mr. Lee Badger, the owner of the Lee’s MarketPlace, a supermarket in Logan, UT, who granted us permission to use his store for the ShopTalk experiments. Special thanks to the very patient department secretaries Genie Hanson and Vicki Anderson who answered all my questions, and to Myra Cook who helped with the final editing of this document.

This work was funded by two Community University Research Initiative (CURI) grants from the State of Utah (2004-05 and 2005-06), NSF Grant IIS-0346880, and NEI/NIH grant 1 R41 EY017516-01A1.

John A. Nicholson
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 CSATL Wayfinding Projects</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1.3 Insight from the Wayfinding Projects</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1.4 Research Goals</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>RELATED WORK</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.1 Introduction</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.2 Grocery Shopping Aids</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.3 Localization and Electronic Travel Aids</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2.4 Landmark Extraction and Autotagging</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>2.5 GIS Data Sharing</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>SHOPTALK: GROCERY SHOPPING FOR THE VISUALLY IMPAIRED</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>3.1 Introduction</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>3.2 Grocery Shopping Task and Spatial Trichotomy</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>3.3 Product Location Assistance for the Visually Impaired</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>3.4 Experiments</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>CRISS FRAMEWORK</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>4.1 Introduction</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>4.2 CRISS: VGI Websites for the Blind</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>4.3 Introduction to the Route Analysis Engine</td>
<td>82</td>
</tr>
<tr>
<td>5</td>
<td>LANDMARK AUTOTAGGING</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>5.1 Introduction</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>5.2 Autotagging Details</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>5.3 Experiments and Results</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>5.4 Summary</td>
<td>114</td>
</tr>
</tbody>
</table>
6 PATH INFEERENCE .............................................................. 115
  6.1 Introduction ............................................................. 115
  6.2 Transformation into a Digraph .................................... 119
  6.3 Inferring Paths ......................................................... 125
  6.4 Path Inference Heuristics .......................................... 129
  6.5 Path Inference Algorithm .......................................... 137
  6.6 Summary ................................................................. 140

7 FUTURE WORK ............................................................. 145
  7.1 ShopTalk ................................................................. 145
  7.2 CRISS and RAE ......................................................... 148

8 CONCLUSION ................................................................. 153

REFERENCES ................................................................. 157

APPENDICES ................................................................. 172
  Appendix A ROUTE DIRECTION TRANSCRIPT .................... 173
  Appendix B NAMED ENTITY SETUP ANNOTATIONS ............... 178
  Appendix C EXAMPLE PATH INFERENCE .............................. 183
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Entries in the Barcode Connectivity Matrix.</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Number of Completed Runs for Each Product Set.</td>
<td>50</td>
</tr>
<tr>
<td>3.3 Demographics of Participants in the Multiple Participant Study.</td>
<td>53</td>
</tr>
<tr>
<td>3.4 Number of Items Scanned for the Experiment.</td>
<td>54</td>
</tr>
<tr>
<td>3.5 Average Values and Weights of the Participants’ NASA TLX Scores.</td>
<td>63</td>
</tr>
<tr>
<td>5.1 Demographics of Route Description Survey Respondents.</td>
<td>103</td>
</tr>
<tr>
<td>5.2 Route Description Set Placement Counts.</td>
<td>104</td>
</tr>
<tr>
<td>5.3 Sentence Counts Per Route Description.</td>
<td>105</td>
</tr>
<tr>
<td>5.4 Word Counts Per Route Description.</td>
<td>105</td>
</tr>
<tr>
<td>5.5 Word Counts Per Route Sentence.</td>
<td>106</td>
</tr>
<tr>
<td>5.6 Landmark Counts Per Route Description.</td>
<td>106</td>
</tr>
<tr>
<td>5.7 Landmark Counts Per Route Sentence.</td>
<td>107</td>
</tr>
<tr>
<td>5.8 Counts of Sentences without Landmark Per Route Description.</td>
<td>107</td>
</tr>
<tr>
<td>5.9 Ratios of Sentences without Landmark Per Route Description.</td>
<td>108</td>
</tr>
<tr>
<td>5.10 Results of Autotagging on Evaluation Route Descriptions.</td>
<td>109</td>
</tr>
<tr>
<td>5.11 Computed Scores for Each Evaluation Set.</td>
<td>110</td>
</tr>
<tr>
<td>5.12 Annotation Types Counts for Each Description Set.</td>
<td>112</td>
</tr>
<tr>
<td>5.13 Entity Annotation Counts.</td>
<td>113</td>
</tr>
<tr>
<td>6.1 Example Landmark Ids for Descriptions A and B.</td>
<td>126</td>
</tr>
<tr>
<td>6.2 Example Route Statement Ids for Descriptions A and B.</td>
<td>126</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Robotic Guide (RG)</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Original WayFinder prototype's hardware.</td>
<td>6</td>
</tr>
<tr>
<td>1.3</td>
<td>WayFinder Wi-Fi collection map.</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>GPS signal drift for two collection points.</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>RoboCart</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Example topological map</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>Route map of the store used in the multi-participant experiment.</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Two shelf barcodes</td>
<td>39</td>
</tr>
<tr>
<td>3.5</td>
<td>Top half of a shelf section</td>
<td>40</td>
</tr>
<tr>
<td>3.6</td>
<td>GUI for BCM data entry</td>
<td>42</td>
</tr>
<tr>
<td>3.7</td>
<td>ShopTalk’s hardware</td>
<td>43</td>
</tr>
<tr>
<td>3.8</td>
<td>Modified barcode scanner resting on a shelf lip.</td>
<td>44</td>
</tr>
<tr>
<td>3.9</td>
<td>Example route for single participant study</td>
<td>49</td>
</tr>
<tr>
<td>3.14</td>
<td>The average run time over repeated runs.</td>
<td>57</td>
</tr>
<tr>
<td>3.15</td>
<td>The average distance walked for an entire run over repeated runs.</td>
<td>58</td>
</tr>
<tr>
<td>3.16</td>
<td>Average number of items scanned over repeated runs.</td>
<td>59</td>
</tr>
<tr>
<td>3.17</td>
<td>Average number of optional LUM instruction requests.</td>
<td>61</td>
</tr>
<tr>
<td>3.18</td>
<td>Sale items stacked in the aisle.</td>
<td>65</td>
</tr>
<tr>
<td>4.1</td>
<td>Route used to compare route descriptions.</td>
<td>68</td>
</tr>
<tr>
<td>4.2</td>
<td>Google Maps’ standard user interface.</td>
<td>72</td>
</tr>
<tr>
<td>4.3</td>
<td>Google Maps’ simplified user interface for screen readers</td>
<td>72</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 Introduction

For a person who has a visual impairment, having the ability to navigate and be mobile in a given environment without the aid of another person is a sign of personal independence. Such personal independence is important enough that a lack of it can affect mental health. Furthermore, the onset of a visual impairment can result in a person having a reduced ability to navigate and travel independently. The reduction in mobility has been noted to lead to depression in some individuals [6]. The ability to travel independently is also a quality of life issue. A 1991 British survey of almost six hundred adults with visual impairments found that only 51% of those surveyed under the age of 60 had left their house alone and on foot in the week prior to the interview [9]. Older persons who have vision loss may also remain home-bound due to a loss of confidence and reduced feelings of independence that can be associated with decreased vision [99].

Montello and Sas [92] define navigation as consisting of two processes: wayfinding and locomotion. Wayfinding is the process of planning how to get to a desired destination. During the wayfinding process, travelers choose and plan efficient routes, noting important landmarks, that will enable them to reach the destination. When planning, travelers often use aids, such as their internal cognitive map or external aids, including physical maps, route descriptions, or electronic travel aids (ETA). Locomotion is the real-time process of navigation during which a person is moving. It is the physical act of movement, e.g., walking, running, or riding a bus. Locomotion also involves solving problems encountered during movement, such as avoiding obstacles and orienting one’s body to landmarks. Wayfinding and locomotion are not necessarily performed independently [91]. Most travelers use both processes during the act of navigation. For example, when encountering and avoiding a
previously unknown obstacle (locomotion) during execution of a previously-planned route (wayfinding), a modification to the original route may have to be planned (wayfinding) and then executed (locomotion).

Unfortunately, people with visual impairments who wish to navigate independently can have difficulty performing one or both of the navigation processes. Route planning effectiveness may be reduced because some people with visual impairments often perceive the world in terms of routes rather than layout [39]. A person may not be aware that two routes intersect, preventing the planning of a new route that uses the beginning part of one route and the ending part of another route. Locomotion can be affected as well. The ability to travel independently requires that a person can successfully move through both indoor and outdoor environments. Indoor environments and outdoor environments each have their own set of characteristics and obstacles [3] often necessitating different skills and strategies. Because of the difference, orientation and mobility (O&M) skills assessment tests evaluate a person’s navigation skills in both indoor and outdoor settings. O&M lessons generally start indoors and later move to outdoors, gradually increasing the complexity of the situations [35].

Localization, a function of locomotion, is an additional problem for people with visual impairments. Localization, the act of determining one’s location in the environment, is critical during navigation since it is necessary to know where one is in order to perform a specific action along a route such as turning at the correct sidewalk or hall intersection. As Golledge et al. [40:217] point out, people with vision are able to “process data in a continuous, integrative, and gestalt-like manner” whereas people with a visual impairment, particularly those with complete vision loss, “actively search the environment in a piecemeal manner.” A person with vision, for example, is able to localize along a route by taking advantage of large-scale geographic observations. For example, while traveling an outdoor route, a sighted traveler can make observations, such as noticing that the next sidewalk intersection is fifty feet away from his current location and that he is walking towards the main clock tower on campus, which happens to be hundreds of feet away. A traveler with
complete vision loss using a white cane, on the other hand, must physically encounter the same sidewalk intersection to actively sense it with the cane and may not be aware his orientation in relation to the clock tower.

1.2 CSATL Wayfinding Projects

In response to the various navigation problems faced by people with visual impairments, two projects have been developed in Utah State University’s (USU) Computer Science Assistive Technology Laboratory (CSATL). The original project was a robotic guide (RG) that used a autonomous robot to guide visually impaired people through buildings. In response to some issues that were identified with RG, WayFinder, a wearable assistive navigation system, was developed.

1.2.1 Robotic Guide

Gharpure’s RG [36] (see Figure 1.1) was an autonomous robot used to guide people with visual impairments in structured, indoor environments. The user would first enter a destination into the system using a keypad on the handle. The user would then grab hold of the handle at the rear, and RG would travel to the destination with the user following the robot. After entering the destination, the user did not have to perform any cognitive
function other than to hold the handle and follow RG. RG was responsible for all navigation
tasks including obstacle avoidance, path planning and execution, and localization.

RG’s obstacle avoidance was based on sensory input from a laser range finder. Laser
readings were processed using a modified potential fields approach, which allowed RG to
follow halls, avoid obstacles, and turn while moving at a moderate walking speed. Local-
ization, and path planning and execution were handled by a radio-frequency identification-
based (RFID) process in which RFID tags were mounted on the walls and doors of the
building. RG carried an RFID reader that was able to read and identify tags as it traveled
past them. Each RFID tag was encoded with a unique id that, when read by RG, uniquely
identified the robot’s location in its internal map. When the RFID tag for the destination
was located, RG stopped and the user knew that he was at the desired target location.

While RG was successful at guiding visually impaired travelers through two indoor
environments, the USU’s Computer Science department and USU’s Center for Persons
with Disabilities building, and was also successfully extended to become RoboCart [37], a
robotic shopping cart for the blind, the system had several issues that would limit its use in
many environments and situations. First, the system was limited to indoor environments.
Buildings with multiple floors would also be a problem if the different floors were connected
only by stairs. The indoor limitation would prevent RG from being used in situations
wherein a person must travel from one building to another building passing through outdoor
environments. A second issue was the laser and the obstacle avoidance process. The laser
was limited to moving back and forth in a horizontal plane approximately 0.3 meters above
the floor in front of the robot. Although extremely accurate within this plane, any obstacle
above or below this plane would not be detected, e.g., waist-high water fountains protruding
from the wall. Another limitation of the obstacle process was that when an obstacle was
detected, users did not know what type of obstacle was present, only that something was
blocking the robot’s path.

A third issue was the speed at which the robot moved. An early version of RG [61]
used sonar rather than the laser range finder. The system moved at a slower speed than
the average walking speed for a human, which some users reported as an issue with the
device. The laser allowed for faster speeds, but would still slow down at turns and in the
presence of obstacles. When an RFID tag was read, localization was very accurate, though
in order for a tag to be read, the robot had to be close enough to read the tag. The range
for detecting a tag was a radius of 1.5 meters. If the robot had to veer away while avoiding
an obstacle and passed the tag outside of the tag’s signal range, the tag was missed and the
robot was unaware of its location.

What should be noted regarding these issues is that people with visual impairments are
capable of performing these tasks, at least to a degree that allows them to navigate familiar
environments. People regularly travel in both indoor and outdoor environments. Using
canes and guide dogs as well as their own senses, travelers manage to avoid large numbers
of obstacles, both stationary and moving. People walk at the speeds with which they are
comfortable and can detect many landmarks based on their senses and understanding of
the environment.

1.2.2 WayFinder

The WayFinder system was conceived and developed as an ETA for the visually im-
paired. In building WayFinder, one of our goals was to build a device that could address
some of the RG’s shortcomings and take advantage of the skills travelers already possessed.
Rather than a large robot, WayFinder was designed to be a relatively small, wearable de-
vice that would help guide independent travelers with visual impairments in both indoor
and outdoor environments. It would assist a traveler with both the planning task and the
localization subtask of locomotion in order to help people in unfamiliar environments. But
unlike RG which assumed control of all navigation tasks, WayFinder was only responsible
for some of the navigation tasks. For the remaining tasks, the user was assumed to have a
sufficient amount of O&M training which would allow him to perform tasks, such as obsta-
cle avoidance and finding landmarks mentioned by the system, e.g., sidewalk intersections
in outdoor environments and hall intersections in indoor environments.

As shown in Figure 1.2, in an earlier version, the hardware was a combination of
off-the-shelf components and a modified computational unit. The user wore a vest that held the components and allowed the individual to use the system in a hands-free mode when responding to prompts from the system. Since WayFinder was meant to be used in both indoor and outdoor environments, it was fitted with several hardware sensors. The primary outdoor sensor was GPS, and the primary indoor sensor was based on an 802.11 Wi-Fi card. A digital compass was added to help determine orientation in both indoor and outdoor environments. The GPS unit was attached to one of the vest’s shoulder straps, and the digital compass was attached to the opposite shoulder strap. An enclosure, which held the computational unit and the Wi-Fi card, was mounted to the front of the vest. A numeric keypad was attached to the front of the enclosure, which allowed the user to respond to verbal prompts from the system. Since all instructions from the system were given to the user as computer-generated speech, the system had a small earphone worn by the user.

Experiments with the WayFinder system yielded some initial results [64, 65, 95, 96]. However, as worked progressed it became clear that WayFinder, like RG, had problems that would limit its success. First, the strategies used to collect signal data were not scalable to large areas, even areas which are relatively small, geographically speaking, such as university
The second problem was that the system’s accuracy was not acceptable for guiding the visually impaired. In order for an ETA to be usable, it must provide the traveler consistent and accurate information, and the system must also be capable of being deployed in a reasonable amount of time and at a reasonable cost.

The collection strategy used for both GPS and Wi-Fi in WayFinder involved collecting locations’ signal fingerprints. The idea for using signal fingerprints for localization was inspired by other Wi-Fi-based localization systems [2, 68, 73]. Wi-Fi signal readings were taken at multiple locations on the halls of USU’s Computer Science department. Figure 1.3 shows the 12 collection points and the 5 locations for the access points. The list of available wireless access points and the strength of each signal from each access point were recorded at each location. As multiple readings were collected over time, a signal-based fingerprint was built up with the idea that a collection of signals can be used to identify a specific location. Since Wi-Fi signals can be noisy and inconsistent, WayFinder used the common approach of applying deterministic and probabilistic methods to help reduce uncertainty when localizing the user [73].

GPS is also prone to noise, which is reflected in the issue of signal drift. Even when a GPS receiver remains stationary, positional readings will change over multiple readings. These errors are produced as the GPS signal passes through the atmosphere and can be a result of nearby objects such as buildings and the leaves in trees [101]. Figure 1.4 shows an example of GPS signal drift for two different collection locations used in the WayFinder experiments. The locations are both on the Quad, a large, grassy area on the USU campus.
One hour of data was collected at each point. The collection period was spread out over six days, and ten minutes of data was collected each day, resulting in approximately 3,600 latitude/longitude readings for each location. As can be seen, the location on the left has much less drift. The location was in the center of the Quad and had no obstructions nearby. The location on the right was collected at the edge of the Quad under a leafy tree and approximately 15 feet from a large building. These obstacles obstructed the GPS signal and caused a wider variation in the readings.

In order to overcome the GPS signal drift, WayFinder used a fingerprinting approach for GPS similar to the Wi-Fi fingerprinting approach. Ten minutes of GPS signals were collected over six days at strategic locations several meters before critical landmarks such as sidewalk intersections. From these readings, a set of standard deviation ellipses were computed for each collection point. The ellipses were used to determine when to give instructions to users. As the user walked a route, the GPS unit would continuously collect data. Whenever the latitude longitude reading was located inside one of the standard deviation ellipses, the location of the user was known and the system gave the next instruction for the route. By placing the collection points before turns, instructions could be given to
the traveler before arriving at the location, giving them time to prepare their locomotion strategy.

A major issue with both fingerprinting collection approaches was the time needed to collect the data in relation to the accuracy achieved. The outdoor data with the GPS was collected over several days in order to capture more of the satellites’ positional changes. The assumption was that since the GPS satellites are constantly moving, instead of collecting the data all at once, collect the data over several days. Indoors, the same approach was taken with Wi-Fi signals being collected over the period of a month. This was due to the noise that the Wi-Fi signals exhibited. It was thought that a more accurate fingerprint could be collected by spreading out the collection over a period of time. Of course, the direct consequence was that the collection time took days, even for one collection point.

While the accuracy of the GPS was acceptable, the accuracy of the Wi-Fi localization was less so. Indoors, Wi-Fi signal fingerprints proved to be not only time consuming but not as accurate as hoped. Wi-Fi signals were not sufficiently reliable for localization of a blind traveler. We discovered, as did the researchers on RADAR [2], that even the orientation of the person affected the signal strength of the Wi-Fi signals. As it turns out, the signal is absorbed by water, of which the human body is largely composed, so placing the Wi-Fi card in close proximity to the person’s body caused a direct effect of the signal. Additionally, interference from materials in the building and mobile devices, such as cell phones which work on the same frequency as Wi-Fi led to accuracy issues.

1.3 Insight from the Wayfinding Projects

As will be seen in the Related Work chapter (Chapter 2), WayFinder and RG joined a long list of devices that have attempted to address the need for indoor and outdoor navigation assistance in unfamiliar environments. Both commercial and research systems have been developed using various technologies including GPS [105, 107], Wi-Fi-based localization [68], and infrared beacons [17]. Unfortunately, the adoption rate for these devices in the blind community remains low. There are multiple reasons for this lack of adoption. First, the commercial devices tend to be expensive. For example, the software and GPS-based
guidance system Sendero GPS [107], intended to be used by an individual, costs $1,495, a price which does not include the mobile computer on which to run the software. Other navigation systems do not achieve localization accuracies that would be useful for a blind person in many situations. Place Lab [68], for example, achieves a median location error of 15 to 20 meters. Other systems [17, 76, 104] require a device or sensor to be installed in the environment. The problems are that each device must often have a power source, or the systems must be calibrated and maintained. These systems do not scale to large-scale environments, such as college campuses, where navigational assistance would be needed over a large area. A final issue is that many systems only address either indoor or outdoor navigation, but not both. A navigational assistance device should ideally address both environments, allowing a person to seamlessly move from one environment to the over without the need to change their ETA.

Another issue is that almost all assistive navigation devices, including both RG and WayFinder, take the “trust me - you are here” approach. In general, they take a reading from their sensor set, compute a location on a map, and, based on the user’s destination, instruct the user where to move next. Unfortunately, due to noise in the environment, signals can be noisy or missing, resulting in incorrect location computation. Garmin [33], for example, reports that its GPS units are accurate to within 15 meters. This amount of error may be acceptable for a sighted person who can make a visual distinction between where the device says the user is standing and where he is actually standing. For a person who is visually impaired, especially those who have complete vision loss, an error of this amount reduces the usefulness of the device. If a person who is visually impaired is continually given an inaccurate location, they may at best simply stop using the navigation device or at worst become disoriented and lost.

One sensor that previous systems have not taken into account is the navigator himself, or rather the navigator’s brain. Many people who have a visual impairment receive extensive O&M training. During training, individuals learn valuable skills that enable them to safely and successfully navigate many environments independently, both indoors and outdoors [6].
They learn, for example, to perform actions, such as following sidewalks, detecting obstacles and landmarks, and crossing streets. They also learn techniques for remaining oriented as they move around the inside or buildings or outside on sidewalks and streets. Experienced travelers can even handle some unfamiliar environments when given sufficient verbal assistance. There is some research evidence [34] that people with visual impairments share route descriptions and can verbally guide each other over cell phones. In these situations, only verbal route descriptions are used to guide the traveler from one location to another location; the only technology used is the cell phone which simply maintains an audio connection between the two individuals.

1.4 Research Goals

This work proposes taking advantage of the previously unused sensor, the independent navigator, by viewing the user as an integral and active part of the navigation system. Research [102] shows that the visually impaired prepare more for travel, make more decisions, and use more information than sighted travelers. Therefore, instead of designing yet another “you are here” type of system, this work proposes to design systems that provide more detailed and user appropriate levels of information. In this case, the more appropriate information is natural language route descriptions that describe how to travel a route. The hypothesis is that if a route is described with the sufficient and appropriate amount of detail, a blind person can use his everyday navigation skills and abilities to successfully follow the route without any wearable sensors or sensors embedded in the environment. The systems described in this paper present natural language route descriptions of quality similar to those given by one visually impaired person to another visually impaired person over a cell phone.

This thesis has four main hypotheses that are addressed by two systems:

1. A navigational assistance system for the blind can leverage the skills and abilities of the visually impaired, and does not necessarily need complex sensors embedded in the environment to succeed.
2. Verbal route descriptions are adequate for guiding a person with visual impairments when shopping in a supermarket for products on shelves located in aisles.

3. Information extraction techniques can be used to extract landmarks from natural language route descriptions.

4. New, natural language route descriptions can be inferred from a set of landmarks and a set of natural language route descriptions whose statements have been tagged with landmarks from the landmark set.

The first hypothesis states that external sensors are not necessarily required in an ETA targeted towards blind navigators. Instead of requiring that a user carry or wear sensors that read signals from the environment, an ETA can be built and designed around the assumption that the user is an independent blind navigator who is a trained and experienced navigator capable of navigating the world and performing most navigation tasks. Rather than a complex system centered around sensors, build a system centered around route directions that describe unknown environments in sufficient detail so as to allow the independent navigator to understand the environment and use his everyday navigation skills.

The second hypothesis builds on the first hypothesis by targeting a specific environment, a modern supermarket, and the specific task of finding individual items on the shelves located in aisles of the store. A grocery store provides a structured environment for testing route instruction guidance, and grocery shopping is also a task many visually impaired people are unable to do independently. The first two hypotheses are addressed by ShopTalk, a system that helps blind individuals independently shop for items located in aisles and on shelves in a grocery store using verbal route descriptions. ShopTalk is covered in Chapter 3.

The third and fourth hypotheses work together in order to expand the results of ShopTalk, moving the concept of using verbal route descriptions to guide a shopper in a grocery store without any additional external sensors to using route descriptions for guidance in any area to be traveled by blind navigators. One source for route descriptions targeted to blind travelers is other blind travelers already familiar with the target area. Chapter 4
describes the Community Route Information Sharing System (CRISS) framework, which is a guide for building a collaborative website where groups of visually impaired travelers can write, edit, and share collections of route descriptions. CRISS is used as a starting point for addressing the third and fourth hypotheses.

The multiple route descriptions collected for CRISS are written in natural language, as that is a natural format for humans. However, natural language route descriptions are unstructured data, which is difficult for a computer to process. The third hypothesis of the thesis, covered in Chapter 5, provides a means of transforming unstructured natural language text into a structure format. Using information extraction techniques, landmarks in sentences can be automatically located and the route description can be transformed into a structure that is amenable to processing by well-known graph algorithms.

The final thesis hypothesis shows how a set of structured natural language route descriptions can be processed to find new route descriptions not originally in the set of user-provided route descriptions. This hypothesis relies on the structure created by the third hypothesis. The set of route descriptions is transformed into a directed graph from which a new natural language description is created. This hypothesis is addressed in Chapter 6. Together the third and fourth hypotheses form the basis for the Route Analysis Engine (RAE), an information extraction-based system that analyzes natural language route descriptions and is the intelligence behind websites built using the CRISS framework.
CHAPTER 2
RELATED WORK

2.1 Introduction

This chapter discusses systems and commercial products that can aid visually-impaired people. The research and technologies discussed are divided into four sections. The first section discusses grocery shopping aids for the visually impaired. The second section covers general navigation aids that can help guide people with visual impairments in both indoor and outdoor environments. The third section discusses tools that have been used for landmark extraction in texts and for autotagging. The last section briefly mentions tools for sharing route information among communities of users.

2.2 Grocery Shopping Aids

There are a limited number of solutions that can help a visually-impaired shopper work towards shopping independently. As stated before, there are two sub-problems involved in shopping: product localization (where is the product in the store) and product identification (what is this product I’m holding). Commercial devices that can provide assistive shopping primarily address the product identification problem.

Not all people with a vision impairment have total vision loss. People who still retain some level of partial vision may be able to successfully navigate large stores and find the general areas of desired products without any assistive device. But due to vision problems, they may still require assistance identifying products and reading information on the product labels. For these shoppers, a hand-held magnifying glass is a low-cost and simple tool. There are also electronic magnifiers on the market, e.g., Looky [110] and OPAL [30], that have advantages over the traditional magnifying glass. These advantages include features,
such as zoom, image enhancement, and contrast adjustment. The disadvantage of the electronic magnifiers is that they require power and are significantly more expensive than a standard magnifying glass.

Furthermore, magnifiers do not meet the needs of all shoppers with visual impairments. As a result, various systems have been developed to provide product identification assistance. One example is the i.d. mate OMNI [24], a commercial product from En-Vision America, which is a standalone talking barcode scanner. It has an on-board database of 2.7 million North American universal product code (UPC) barcodes and descriptions of the items. For some products, the database contains additional information beyond the basic description such as product ingredients, nutrition information, package size, and warnings. If the user finds an item with a barcode that is not in the system’s database, the i.d. mate OMNI allows the user to add the new barcode and associated information. Strictly speaking, the system does not provide product localization support. However, the system does have the capability for the user to record audio memos. This feature could be used for recording instructions on how to find products in a store, but this would not scale well for large numbers of products.

Another approach to identifying items is based on extracting text found in images. The primary example of this technology is knfb Reading Technology Inc.’s software called kReader Mobile [57]. The software runs on Nokia N82 cell phones. The user takes pictures of objects with the phone’s camera. The system then performs OCR on the image and uses text-to-speech to read the extracted text to the user. While not explicitly designed for assistive shopping, the device can be used in a grocery store setting as a product identification tool. Users could identify products through processing images of the desired products. As with the i.d. mate OMNI, this system does not address the problem of finding the target product in the first place; the shopper must have some other method to find the product’s location.

In addition to commercial devices that can be adapted to shopping situations, there have been several research projects in recent years that have specifically addressed the issue
of designing an assistive shopping device for people with visual impairments. RoboCart [37] from Utah State University was the project that preceded ShopTalk. RoboCart was similar to ShopTalk in that it allowed people with visual impairments to independently shop for shelved items in a grocery store. The major difference between RoboCart and ShopTalk was that instead of guiding the shoppers through the locomotor space with verbal instructions and allowing the shoppers to use their own O&M skills to accomplish the navigation tasks, RoboCart used a specially designed shopping cart with a robotic base. The shopper would enter the product that he wanted to find in the system, grab hold of the cart’s handle, and the robotic base would travel to the general area of the product pulling the shopper along. In the locomotor space, the shopper was not required to make any decisions. The robot determined its location using a combination of laser-based Monte Carlo Markov localization (MCL) [29] and RFID tags placed in mats which are on the floor. Whenever the robotic cart passed over an RFID-embedded mat, it updated the robot’s location in the MCL process since the RFID was a well-known location and detecting the RFIDs could be done accurately. One issue with RoboCart was not technical, but cost. Given the slim profit margins of grocery stores, a cart with a robotic base would be too expensive. Another issue was that many of the shoppers were able to walk faster than the system was allowed to go.

Trinetra [70] from Carnegie Mellon University (CMU) is an assistive shopping device targeted towards small stores such as CMU’s on-campus convenience store. The shopper carries a Bluetooth-enabled cell phone, a Bluetooth headphone, and a Bluetooth barcode scanner. When shopping, the user scans the product’s UPC barcode. The application first checks a data cache on the phone to see if the product has recently been scanned. If the barcode is in the cache, the application provides a description of the product to the user. If the scanned barcode is not in the cache, a remote server is contacted to provide the product’s information. Because the store used in tests was small and consisted only of two main aisles, the system does not provide directions to the user on how to effectively manage the locomotor and search spaces. The implicit assumption is that the shopper is sufficiently familiar with the store to locate target products. If the shopper has a problem
finding an item, it is assumed that the cashier is only a few feet away and available to help. This assumption, while valid for Trinetra’s target store, does not hold in large, modern supermarkets.

Grozi [88], which was developed at the University of California, San Diego, uses image recognition to identify individual products. 676 images of the product captured under ideal situations and 11,194 images image found on the web were used to represent 120 products. The ideal images were then used as training data and the web images were used as the test data. The SIFT algorithm [80] was then used to identify products from images captured by a camera on a mobile device. Grozi is primarily focused on the product identification aspect of shopping and does not address how to get to the product’s location.

The iCARE system [58,59] is built on the concept of item-level RFID tags, i.e., each product on the shelf has its own RFID tag. The user wears a Bluetooth RFID reader on his wrist that communicates with a PDA. When the reader reads an RFID tag embedded in a project’s packaging, the PDA communicates through Wi-Fi to a database server and can then identify the product. The system does not address how to guide the shopper to the target item’s haptic space; it assumes the user is already in the vicinity of the product.

Item-level RFID tagging may provide benefits over barcodes for visually impaired shoppers. The user does not have to locate the RFID tag’s exact position in order for it to be read as is necessary with barcodes. The tag just needs to be in the vicinity of the RFID reader. The reason that ShopTalk uses barcodes rather than RFID is that item-level RFID tagging has largely been limited to pilot studies and has not been widely adopted. Some consumers are reluctant to use RFID-tagged products due to privacy concerns [69]. Once a person is carrying an item with an RFID tag, there is a perception that they can be tracked inside and outside the store. Another issue is that RFID tags are still too expensive to support widespread item-level tagging [124]. Technology limitations also have hampered adoption. Some item and packaging material, such as metal cans, may affect the ability of readers to reliably read the tag [111]. It is unknown when item-level RFID tagging will be implemented on a wide-spread level, but using barcodes as a localization method can
be done now. If and when RFID item-level tagging is adopted, ShopTalk can easily be migrated to use the alternative sensor.

The patented system described in [13] is similar to ShopTalk in that it is based on using barcodes to find items in the store. The difference is that this system requires its own placement of barcodes rather than co-opting the existing shelf barcodes used by the store inventory system. The system first places barcodes representing each product at the aisle end. Another barcode is placed on the edge of a shelf, representing the shelf. The UPC barcodes on the physical product need to be scanned in order to locate and identify the target product. From a grocery shopping perspective, the end-of-aisle labels would not be sufficient for most aisles in a modern grocery store. In the patent description, eight items are shown on one side of the aisle whereas in a real-world grocery store, one side of an aisle can have hundreds of products. Another issue is that finding the actual item requires picking up an item to scan the UPC code. Scanning inventory barcodes on the shelf is quicker and requires no manipulation of any object. Shelf locations are also encoded as distances rather than relative positions.

Except for RoboCart and the system described in [13], these products and systems primarily address the issue of product identification, not the product search task. If the shopper is able to find the approximate location of the target product, primarily in the haptic space, after limited searching by the user, these systems can help to ensure that the product picked up is the correct product. However, these systems would be tedious to use, or even unusable, for many shoppers since the systems do not provide product maps or guidance in the locomotor space, and only limited guidance in the search and haptic spaces. If the shopper needs any sort of assistance for navigation, these devices will not meet users’ needs.

2.3 Localization and Electronic Travel Aids

CRISS, with its RAE subsystem, is an assistive navigation system for indoor and outdoor environments. ShopTalk, when providing instructions to a shopper on how to navigate through the locomotor and search spaces of the grocery store, also functions as an
assistive navigational device, albeit for a specific type of structured indoor environment. In both of the systems, the navigator’s senses are the primary method for gathering information from the environment. Other systems have been designed that take the more traditional approach of using external hardware sensors to provide localization and navigation support for indoor environments, outdoor environments, or both.

Talking Lights [76] is a commercial system that works with existing light fixtures in indoor environments. Devices are installed in the light fixtures that modulate the light at different frequencies. This technology enables the transmission of data that can be captured and decoded by optical receivers carried by a user. Talking Lights has been tested as a wayfinding aid [72] for the visually impaired. Eight participants were tested on four routes in a building in which Talking Lights had been installed in a series of fluorescent lights. The data emitted by the system were route directions. The researchers did not find a difference between the time it took the participants to complete the routes when following the Talking Lights-based instructions and when following verbal route instructions provided by the researchers. Because of its dependence on the lighting system, Talking Lights is primarily an indoor-based system.

Talking Signs [17], developed by Smith-Kettlewell Eye Research Institute, is similar to Talking Lights. Instead of modifications to light fixtures, Talking Signs requires that special infrared transmitters are placed at strategic locations in the environment. The transmitters encode and transmit recordings of human speech. The system has an element of directionality in that the infrared signal spreads out like a cone from the transmitter; this helps to reduce signal overlap in areas with more than one transmitter. The transmitter’s range can be set from 3 meters to 18 meters. The user carries a receiver that detects the transmissions and decodes the speech. The receiver can distinguish multiple transmitters based on signal strength. The user must actively move the receiver in order to find the transmitter signals. Once the desired signal is found, the navigator can walk towards the signal. Talking Signs can be used outdoors, as well as indoors, as long as there is no competing infrared source. In outdoor environments, sunlight can reduce the effective range
of the device.

The Tactile Map Automated Production (TMAP) system [89], also developed at Smith-Kettlewell, is an example of a system that can produce a tactile map. A tactile map is a map for people with visual impairments in which the visual information is supplemented or replaced with raised features that can be explored through touch. In general, the design for tactile maps is simpler than their visual counterparts’ design in order to ensure that the map can be read [4]. TMAP’s data source is the US Census Bureau’s Topologically Integrated Geographic Encoding and Referencing System (TIGER)/Line data [117], which contains features, such as streets, roads, and rivers, and geographic boundaries. TMAP has been extended to work the Talking Tactile Tablet [89], which provides a more interactive approach than a printed tactile map and a real-time, wayfinding solution based on tactile maps. The issue with TMAP is not with the technology but with the data source. TIGER is meant to support census reporting, not assistive navigation for the blind. The level of detail may not be sufficient in some areas for a person with a visual impairment to follow a route safely. Because of the source data, TMAP is primarily for outdoor environments, but tactile maps are also designed for indoor environments by commercial companies [25].

The system described in [15] attaches simple, colorful targets with large barcodes to indoor locations and then uses computer vision applications running on a cell phone to locate the signs and guide the navigator. When the color target is detected, it signals the presence of the barcode. The barcode is a simplified 1-D barcode that is large enough to be decoded from a distance. The barcode can represent different items, depending on the application. During experiments, the barcode represented an index into a database that provides location information. The color targets could be located from 2.5 to 11 meters away depending on target size, amount of illumination, and the phone’s camera resolution.

A technology used in a similar manner to infrared-based localization is ultrasound. The Bat System [45] uses ultrasonic devices to provide localization and tracking services. Cricket [104] and Cicada [56], both localization systems, also use ultrasonic devices as well as radio frequency (RF). Drishti [105], an indoor and outdoor navigation system for the blind,
integrates an ultrasonic localization system from the Australian company Hexamite [46] to help with indoor localization. In general, these systems require that a network or grid of beacons or transmitters be installed in the ceiling or on the walls. The user then carries a receiver which is used to triangulate its position. Positioning for these types of systems has been accurate for the purposes of a wayfinding system. The Bat System, for example, uses time-of-flight of the ultrasonic signals and is accurate to within 9cm for 95% of its readings. Cricket also uses time-of-flight and has achieved an accuracy of 4x4 feet cells. Cicada is able to provide locations within 5cm average deviation.

Radio-frequency identification (RFID) localization-based systems are typically used one of two ways. In one version of this type of system, RFID tags are placed at strategic locations in the environment, and the traveler carries an RFID reader. In the other version, the traveler carries an RFID tag and multiple RFID readers are placed around the environment. In both cases, whenever an RFID reader detects an RFID tag, the traveler’s location is then known to be within some distance of the fixed location of the static reader or static tag. The effective distances for reading RFID tags range from several centimeters to tens of meters depending on the specific type of RFID tag and reader used, thus affecting the accuracy of the localization system.

An example of an RFID-based system is LANDMARC [94], which uses stationary RFID readers. Stationary reference RFID tags are placed around the building, thus providing a means of location calibration. Location is calculated using a mapping from power levels of received signals to RSSI for both the reference tags and the tags on the mobile users. In experiments, four readers and one reference tag per square meter located objects within 1 to 2 meters. Another system, RadioVirgilio/SesamoNet [21], places a matrix of small, cheap, passive RFID tags in a carpet and then a user carries a cane with an RFID reader embedded in the tip. RoboCart [37], mentioned earlier also used RFID tags as part of its localization method. The difference here was that the RFID reader was carried by the robot rather than the human shopper.

Systems that receive and analyze radio frequencies are another type of localization
system. The SpotON system [47] developed at the University of Washington used tags that randomly transmit beacon signals. When other tags detect a beacon’s signal, they use the received signal strength information (RSSI) to estimate distance to the transmitter.

Other systems that use radio signals often use frequencies and protocols that are standard-based, enabling the use of off-the-shelf hardware. For example, as mentioned in the introduction chapter, localization systems based on Wi-Fi have been developed. Wi-Fi has been used in systems, such as the one mentioned in [49] in which the received signal strength from standard Wi-Fi access points is measured and then used to localize the user to a point inside a building. The RADAR system [2] from Microsoft Research is designed to locate and track users inside buildings. It processes the Wi-Fi signal strengths received by a standard network cards received from 802.11 base stations, using a signal propagation modeling method to calculate the mobile users’ positions and is capable of achieving a median error distance of 2 to 3 meters. The Place Lab-based system [68] described in [120] performs sensor fusion on GSM, Wi-Fi signals, and data from an accelerometer and provides location information with an average accuracy of 20 to 30 meters. Using Wi-Fi fingerprinting like RADAR can improve the resolution to 1 to 2 meters. Commercial products now exist that provide Wi-Fi-based localization services. For example, the Ekahau RTLS (Real Time Location System) sold by Ekahau [23] uses Wi-Fi, with the company reporting an expected accuracy of 1 to 3 meters. Their system collects RSSI values from Ekahau tags worn by users, which are processed by a server to determine the tags’ locations.

Similar to the Wi-Fi systems are those based on Bluetooth with the main difference being that the Bluetooth range is far lower, more akin to RFID-based systems. The system described in [121] uses a Bayesian-based approach on the RSSI received from Bluetooth dongles. Essentially, a variation on the triangulation approach, it achieved an accuracy of 2 meters with a standard deviation of 1.2 meters when using three service points. One of the newest wireless standard to be exploited for localization is the ZigBee standard [128]. In [112], the RSSI from nodes on a ZigBee-based network is used for localization achieving an accuracy of 1.5 to 2 meters.
In outdoor environments, GPS is the most common technology used due to its ubiquitousness. Sendero GPS [107] and Trekker [50] are both commercial GPS products specifically designed for independent navigators with visual impairments. In outdoor environments, the previously mentioned Drishti [105] also uses GPS. WayFinder Access [118] is a commercial GPS cellphone product that provides text-to-speech navigation for the visually impaired. Loadstone GPS [77] is a GPS-based, open source, navigation tool for Symbian S60-based cell phones.

One disadvantage of these systems is the overhead required to manage these systems and instrument the environment. Tags and beacons need to be installed at the best locations, and, depending on the system, may be perceived as negatively affecting the appearance of the store. Several systems require multiple devices that require power. If battery powered, devices need to be monitored on a periodic basis. If the devices are not battery powered but power is still required, extra power infrastructure may need to be installed. Additionally, calibration may be required for most of these systems in order to achieve the highest accuracy.

2.4 Landmark Extraction and Autotagging

Other research has investigated extracting landmarks from text. In the majority of this research, the targeted landmarks are locations or addresses of locations mentioned in text, often web pages. Although information extraction is one of the technologies used to accomplish this task, it is not the only technique used for landmark extraction. In all cases, these systems focus on extracting much larger landmarks than those that RAE extracts. The landmarks in these systems are typically street addresses or building names, or even larger geographic areas such as cities.

In Loos and Biemann [79], the target landmark set consists of address information for places like restaurants, cinemas, and shops in web pages obtained through Google query result sets. The open source MALLET tagger [86] was trained to extract address information: city, street, house number, zip code, etc. This was not strictly the NE sub-task of IE because the goal was to only extract the address for the target real-world location described
in the web page. In a standard NE task, all addresses found in the web page would have been extracted.

Another group [114] extracted Japanese style addresses from web pages. They first created a dictionary of terms related to addresses from the Japanese Ministry of Postal Services. The terms were then assigned to one or several of the hierarchical levels of prefecture, city, and town. Addresses were extracted when text in web pages contained terms in the dictionary and matched address pattern expressions. The dictionary was then used to process 11,200 web pages and marked 2,598 pages as having spatial data. This dictionary method is similar to the gazetteer look-up in ANNIE. Any word not in the dictionary, or gazetteer is not marked.

In [81], the IE task of extracting geographic locations was bootstrapped with a knowledge base consisting of location data. The knowledge base replaces a flat gazetteer, like the ones supported in ANNIE. The advantage of the knowledge base is that spatial information is part of the data and that information can be used to enhance the extraction process. The location knowledge base is primarily populated with data from NIMA’s GEOnet Names Server [93], but other data sources are used as well. The knowledge base was used in conjunction with ANNIE from GATE [19] to extract geographic locations in 101 online newspaper articles. The extracted locations were compared with locations extracted by MUSE [85], a multi-purpose named entity recognition system. Although MUSE performed slightly better, extracting more entities, the knowledge base approach was able to categorize locations better, i.e., whether the location was a city, a province, a country, etc.

The InfoXtract system [74] addresses the issue of ambiguity in location names. For example, there are 23 cities in the United States with the name “Buffalo.” Geographic names were extracted from news and travel guide web pages using NE. In order to disambiguate the extracted entities, a graph was created with the entities as nodes and edges representing the relationships between the nodes. The graph was then processed by a modified version of Prim’s algorithm to find the minimum spanning tree. The edge weights were calculated based on a predefined weight table.
Tezuka and Tanaka [113] created five measures for evaluating the significance of landmarks mentioned in web documents. The measures included the frequency with which the landmark was mentioned, whether or not landmarks were co-mentioned with other landmarks, whether or not landmarks were mentioned with place names, frequency of landmark mentions in sentences with spatial sentences, as well as mentions of landmarks in deep spatial structures. Spatial sentences are sentences that contain place names and spatial triggers, such as verbs or prepositions, identifying spatial relationships. Deep spatial structures are linguistic structures that could be identified through the information extraction subtasks template relation construction and scenario template production. These measures were evaluated by extracting landmarks from 157,297 web pages related to Kyoto, Japan. The landmarks to be evaluated were chosen from a GIS database that identified landmarks in Kyoto. The evaluation compared the significant landmarks identified in the web pages against landmarks listed by humans as important landmarks in Kyoto.

The research in [109] is concerned with assigning geographical scope to web pages so that spatial queries can be performed on sets of web pages in a manner similar to queries performed on traditional GIS data. As part of the procedure of assigning geographic scope, the system performs a named entity recognition step that extracts place names from each web page. The traditional NE approach is extended to handle multilingual geographical references. Ambiguities related to place names are resolved using language patterns, features, and relationship in the text to other place names. After the NE step, documents are passed to another process that assigns the final geographic scope to the document. This step relies on a geographical ontology that models relationships between different geographic regions. This model also helps resolve any remaining ambiguities not handled by the NE step. The document is processed by a variation PageRank algorithm [100] using the extracted place names to assign it to the most likely node in the geographic ontology.

Landmarks have been extracted not only for human use, but for robots as well. In [71], natural language instructions are used to give instructions to robots. The system converts verbally spoken instructions down to sensory-primitive actions. Because the correct map-
ping of natural language must be made to a specific set of sensory-motor primitives, this
system must handle the topological, causal, control, and sensory levels as defined in the
SSH [60]. RAE, and ShopTalk as well, assume that the person being guided does not need
such a low-level mapping; the systems assume the person knows how to do things like avoid
obstacles, find doors, follow sidewalks, etc. RAE is only concerned with information at the
higher, topological level.

The system described in [127] found and extracted routes in HTML pages. The system
extracts sentences from the HTML documents containing route directions using natural
language knowledge and HTML tag information. The sentences were then classified as one
of four classes: destination, origin, instruction, or other. Destination and origin are the
starting and ending points of the route description. An instruction represents one sentence
in a set of sentences. Sentences classified as other are sentences not related to the actual
directions. Sample documents with route directions were identified by an Internet search.
Sentences are classified based on several features including bag-of-words, HTML tags in the
original HTML encoding of the sentence, and domain specific features related to directions
such as verbs, e.g., “turn,” and distances, e.g., “miles.” Sentences were then classified using
four different models: naive Bayes, maximum entropy, conditional random fields (CRF)
and maximum entropy Markov models (MEMM), were evaluated on 10,000 human tagged
sentences from 100 HTML documents, with CRF and MEMM performing the best. There is
no indication that individual landmarks other than the destination and origin are identified.

In CRISS, landmarks are considered tags for route statements, and RAE is used to
autotag the route statements, i.e., assign landmarks automatically. Tags are also a common
feature found on many websites used to annotate text, images, and so. These tags are often
entered by users, but systems have also been developed to provide autotagging features.
The main difference in these cases is that the tags in the other systems are not as focused
as the tags in CRISS and RAE. In CRISS and RAE, the tags are only landmarks, but in
these other systems, tags can be any term or phrase chosen by the users.

AutoTag [90] suggests tags for blog posts. Tags are automatically selected based on
tags found in other blog posts that are similar to the user’s blog post. Similar blog posts are found through the highest ranking results of a search. Tags are then selected on a frequency count based on usage in the search result blog posts. During evaluation, users retained four to six tags recommended by AutoTag.

Fujimura, Fujimura, and Okuda [12] describe another system for automatically recommending tags for blog entries. They describe a k-nearest neighbor approach to recommending tags for blog entries. A potential list of tags is first formed by performing a related document search on previously tagged documents. Tags are then selected from the result set based on the probability of how often the tag is used. The system also addresses the issue of reducing the number of potential tags by reducing the number of tags that are synonyms; tags are considered synonyms if they have a high score based on the cosine-based similarity measure.

Brooks and Montanez [8] are primarily concerned with how effective tags are at classifying blog entries. However, they also extract the three words with the top term frequency, inverse document frequency (TDIDF) score in blog entries; the top three words become the document’s autotags. Although extracted automatically, the documents with shared autotags were considered more similar than documents with shared user-provided tags when clusters of similar documents were measured using the average pairwise cosine similarity.

2.5 GIS Data Sharing

CRISS and RAE were designed on the idea that users can create and share route descriptions in a manner similar to wikis. Existing navigation tools also provide mechanisms that allow users to share navigation and localization data. However, since these existing tools are primarily GPS based, the information is restricted to outdoor locations. Most of the shared information consists of points of interest, and not entire route descriptions. Routes are not typically shared due to the fact that they are usually planned and automatically generated by the system based on its set of available landmarks.

The developers of the open source Loadstone GPS [77] maintain the website Loadstone PointShare [78]. The site allows users to upload the name, latitude, and longitude of new
points of interest, as well as download data previously uploaded by other users. The site only supports sharing of points of interest. It does not have features that allow users to share routes.

Sendero GPS [107] also maintains the MySendero [106] website. Users can contribute points of interest and can download other users’ contributions. This is similar to the Loadstone PointShare website.

An additional feature of previously mentioned WayFinder Access [118] is that the system provides access to the website MyWayfinder [119]. This site allows users to download and share trips, routes, and destinations. The site also has support for points of interests like Loadstone PointShare and MySendero. Unlike those websites, it also allows users to email routes and maps to other users.
CHAPTER 3
SHOPTALK: GROCERY SHOPPING FOR THE VISUALLY IMPAIRED

3.1 Introduction

The list of the most functionally challenging environments for individuals with visual impairments is topped by shopping complexes [102]. This difficulty can be understood given that a typical modern supermarket stocks an average of 46,852 products and has a median store size of 46,755 square feet [28].

Many people with visual impairments do not shop independently. They receive assistance from a friend, a relative, an agency volunteer, a store employee, or a hired assistant [55, 62]. Depending on the assistant’s availability, the shopper may need to postpone the shopping trip. Even the ability to get to the store independently does not preclude mishaps when attempting to go shopping. A university student, who has total vision loss and who participated in the multi-participant study reported later, related a personal anecdote in which he was required to wait for over 15 minutes at a local supermarket before an employee was able to assist him. Although he is a highly skilled independent traveler, capable of independently taking the bus and walking around Logan, Utah, and the USU campus, he is unable to shop independently due to his lack of sight. His account agrees with some stories in popular press on the shopping experiences of individuals with visual impairments [10].

In addition to having an assistant help at the store, another grocery shopping option for people with visual impairments is to use a home delivery service, such as the Internet-based PeaPod [103] or those provided by some brick-and-mortar grocery stores. Shopping lists are usually provided over the phone or through online websites. The two main disadvantages
of delivery services are that they are not universally available and that they require the shopper to schedule and wait for the delivery. Although these services are useful, personal independence is reduced and spontaneous shopping is not possible in this context. While a useful option when available, they do not fulfill all shopping needs.

3.1.1 RoboCart

RoboCart [36, 37] (see Figure 3.1) was a robotic shopping assistant for the visually impaired. An extension of RG [36], it guided the shopper to the approximate location of a desired product. The robot would then stop and inform the shopper he was near the desired product. The shopper would then take a wireless barcode scanner and scan the shelf barcodes, which are barcodes used by the store’s inventory system and are attached to the front shelf edge directly below the product. Once the product was found, the shopper placed the item in the basket that was carried by RoboCart. When the shopper was ready, RoboCart guided him to the next product on the shopping list, or if all items on the list had been found, RoboCart would guide the shopper to the cashier lane.

An important modification made to RG for RoboCart was that floor mats with small RFID tags embedded in them replaced the large RFID tags that had hung on the wall
during testing with RG. The RFID antenna that was originally situated waist high was moved to the front of RoboCart and focused on the floor so that the RFID-embedded mats could be read. Another modification was that a camera was added to the front of RG. It was difficult to place the RFID mats in large open spaces because RoboCart would often miss the mats. This was not a major issue for RG since it was deployed in buildings consisting of halls that limited the distance it could veer. The open areas in the grocery stores, on the other hand, were large enough that when RoboCart was required to avoid an obstacle, it was possible that it would veer too far and miss the RFID mats. The camera was used to implement a line-following strategy. Rather than RFID mats in the open areas, blue tape was placed on the white floor connecting aisles and allowing RoboCart to navigate these areas without becoming lost. When obstacles were detected, it was assumed it was a shopper or employee, and RoboCart would ask the person to move; it would not veer from the line. A final modification was that the frame which originally held the RFID antenna was also modified so that it could support a hand basket allowing RoboCart to act as both guide and shopping cart.

Like RG, RoboCart had limitations that would inhibit its adoption. The first issue was a limitation imposed by the research scope. Because the research was more focused on the robot navigation issue, the problem of searching for products on the shelf was not fully addressed. The map of products was limited to only a few items on one shelf. When users scanned the products, the users were told which shelf to scan beforehand. They did not perform a complete search that assumed the target product could be on any shelf and at any position. Another issue was that the robot did not stop directly in front of the desired product, requiring the user to maneuver around the robot in order to start searching for products. As with RG, RoboCart performed all navigation tasks; the user entered a destination and then held onto the robot’s handle. This required the user to travel in a manner that he was not accustomed to and did not take advantage of his navigation skills.

RoboCart’s other issues were cost-related. During the experiments, the RFID mats needed to be placed around the store. In a real-world installation, the RFIDs would most
likely be embedded directly in the floor. There would be a cost associated with this installation. Finally, the robot itself was expensive, costing around $10,000, and in a real-world situation would also have to be maintained by store staff. Given that stores already support blind shoppers with store staff and at a far lower cost, albeit at a service level that is unacceptable to many visually impaired, the additional costs of a robotic shopping cart make it unlikely to be adopted.

3.1.2 Focus Group Comments

In order to further provide some background on the problems faced by shoppers with visual impairments, the following discusses comments made during an informal focus group regarding the general shopping experience. The focus group was conducted in March 2007 with six visually-impaired individuals from the Cache Valley area in Utah. All six individuals were independent travelers who held part-time or full-time jobs, used public transportation, and walked independently around their neighborhoods. The focus group was part of a regular monthly meeting of the Logan Chapter of the National Federation of the Blind that meets at USU’s Center for Persons with Disabilities. The theme of the focus group was independent blind shopping in modern supermarkets. The focus group participants were asked four questions:

1. Do you go grocery shopping on your own?

2. How do you find the right products in the store?

3. What are the typical difficulties that you have to overcome in the store?

4. What devices or equipment would you like to have to improve your shopping experience?

In response to the first question, four out of the six individuals gave an affirmative answer. One individual said that he used to go grocery shopping independently before he got married. After he got married, he started shopping with his sighted spouse. Another individual said that she did her shopping only in the company of a sighted friend or relative.
In response to the second question, all six participants said that they used the help of a sighted guide who was either an assigned store staffer or an individual who accompanied them to the store.

When answering the third question, all four individuals who answered affirmatively to the first question reported delays and having to wait for assistance in the store. However, this was not the only difficulty that they reported. Another difficulty was the competence of the store staffers assigned to assist the shoppers. Some staffers, who were assigned to help the shoppers, were not familiar with the store layout or the store products. Consequently, the participants either had to give up searching for the product that they wanted or had to settle for products that were distant substitutes. Two participants reported that they had had assistants who became irritated with long searches for exact products, which caused the participants to stop searching. One participant said that on several occasions she was assigned staffers who could not speak English and could not read the products’ ingredients to her. Another participant said that once she was assigned a cognitively disabled staffer who was able to guide her around the store but was unable to assist her with product searches.

The fourth question was admittedly open ended. The answers included technological ideas such as RFID tags on all products and cellphone cameras that can recognize all products. Other ideas were non-technical such as training staff assistants on how to help visually-impaired shoppers. All participants said that, in addition to shopping for specific items, they would also appreciate the ability to explore the store independently so that they could learn its basic layout.

3.1.3 Research Scope

A previous investigation of independent blind shopping [63] distinguished two types of grocery shopping: small-scale and large-scale. In small-scale shopping, the shopper buys only a few items that can be carried by hand or in a hand basket. Large-scale shopping involves buying more products and typically necessitates the use of a shopping cart. When designing RoboCart [37,67], the ergonomics-for-one framework [87] was used to interview ten
people with visual impairments about their grocery shopping experiences. The subsequent analysis of the interviews identified five subtasks of the generic grocery shopping task: 1) traveling to the store; 2) finding the needed grocery items in the store; 3) getting through a cash register; 4) leaving the store; and 5) getting home.

This research focuses on the second task of finding the required grocery items in the store. The scope of this investigation is limited to shopping for items stocked on the aisle shelves in a typical supermarket. The scope is further restricted to small-scale shopping. Tasks such as shopping for frozen products and produce as well as large-scale shopping are beyond the scope of this investigation. This research also is limited to product localization, i.e., finding a product’s location on the shelves.

This research does not address the problem of product identification, since it is not seen as a major obstacle. Well-managed stores maintain their shelves and the products on the shelves because their business depends on it. Employees regularly inspect the shelves to remove misplaced items, restock the sold items, and move items to the front of the shelf so that items are within easy reach. As long as a shopper picks up the product immediately above the shelf barcode, product identification will not be a major issue in most cases. Many people with visually impairments have some residual vision that can help them identify products. Haptic cues can also be used to distinguish items, e.g., a peanut butter jar can be easily distinguished from a can of corn by touch. In the case of identical containers, either a second scan of the barcode on the product’s label or verification at checkout could be used to resolve product questions. The problem of individual product identification can most likely be solved with a combination of technical solutions, e.g., computer vision, and non-technical solutions that rely partly on the shopper’s intelligence and ability to adapt and partly on the willingness of the store to keep their customers satisfied.

3.2 Grocery Shopping Task and Spatial Trichotomy

The task of searching for grocery items, when performed by a typical sighted shopper with a shopping list, has three basic stages. The first stage begins after the shopper has entered the store and possibly obtained a hand basket or a shopping cart. The shopper
orients himself to the store layout and begins to search for the first item on the shopping list. The current item on the list is referred to as the target product. During the second stage, the shopper walks around the store trying to find the target product. When the target product is found, the shopper places it into his hand-basket or wheeled cart, if one is present, and then proceeds to search for the next item on the list. The second stage is completed when the last item on the shopping list is found. In the third and final stage, the shopper travels from the last item’s location to the cashier, pays for the items, and leaves the store.

The second stage requires that the shopper both locate and identify each individual product on the shopping list. Product localization involves determining where a product is located in the store and then navigating to that location. Product identification is the process of ensuring that the found product is the target product. Product identification is necessary because the target product may not be at its expected location. The product may be sold out, a different item may be in the target product’s location, or the shopper may have picked up an item adjacent to the target product. Product identification is also critical for distinguishing between certain types of products. For example, many canned vegetables can only be differentiated from one another by performing a visual inspection of their label.

When the shopper is looking for products, he is moving through the space within the grocery store. Inspired by the research of Barbara Tversky [116], the different types of space within the shopping task have been categorized in order to better understand the shopping task. Previous research with RoboCart [37, 63] used the standard spatial dichotomy (locomotor vs. haptic) from the blind navigation literature [31]. The locomotor space includes areas of large-scale movement around the grocery store. The haptic space is the space in the immediate vicinity of the shopper’s body, i.e., the space the shopper can reach without locomotion. RoboCart guides the shopper around the grocery store in the locomotor space. When the robot reaches the approximate area of the desired product, the shopper uses a bar code scanner to search for the product. This search involves the shopper
reaching out to products in the haptic space.

In this work, the dichotomy of locomotor and haptic spaces has been extended to better represent the task of searching for individual products in aisles. Although it is possible for a shopper to move through the locomotor space and stop immediately in front of the current target product, placing it in the shopper’s haptic space, it is more likely that a shopper will get to the general location of the target product and then use smaller-scale locomotion in order to fine-tune his location to be directly in front of the target product. The additional space, called the search space, is defined as a small space around the shopper where the shopper performs limited locomotion, narrowing in on the specific location of the target product. Because there is locomotion, the search space overlaps with the locomotor spaces. This trichotomy may be appropriate both for sighted shoppers and visually-impaired shoppers, and draws on the research by Freundschuh and Egenhofer (FE) [31] who provide a comprehensive review of the previous work on categorization of space.

The task of locating a product in the store can be viewed as a process of the shopper moving through the three spaces. In the locomotor space, the shopper travels from his current location to the general area of the target product. This involves such actions as walking to the correct aisle, entering an aisle, and walking to the general area where the shopper expects the product to be. Moving to the product’s expected area does not guarantee that the shopper will be directly in front of the product. Although stores tend to group similar products together in sections, some of the product sections, e.g., canned soups or pasta sections, can be too large for the shopper standing at one end of such a section to physically reach the products at the other end of that section without locomotion.

Once in the vicinity of the product, the shopper shifts from the locomotor space to the search space. In the search space, the amount of locomotion required may be as small as one or two small steps before the shopper can place himself directly within reach of the target product. At other times, such as when searching for a product in large sections of similar products, the amount of locomotion may be a few meters, but still small in comparison to the amount of locomotion and effort required in the locomotor space. One feature of the
search space that is different from the locomotor space is that a small amount of visual or haptic scanning may be required to determine the position of the target product.

When the target product is within the reach of the shopper and requires no additional locomotion, the product is considered to be in the shopper’s haptic space. This space requires no locomotion on the part of the shopper, because the shopper can now physically grasp the target product. There may still be an element of search involved, e.g., scanning the different shelves, in order to find the exact location of the target product.

3.3 Product Location Assistance for the Visually Impaired

The above functional analysis of the shopping task sheds light on the requirements for independent shopping solutions for the people with visual impairments. To guarantee independence, any device must, at the very least, enable the shopper to navigate the store reliably and to search for and retrieve individual products. One way to satisfy these requirements is to instrument the store with various sensors, e.g., Talking Lights [76] or RFID tags [37, 94], and give the shopper a signal receiver that provides directions as a function of the identities of the sensors detected in the environment. The same instrumentation approach can be carried over to the task of product search and retrieval. For example, one can assume that every item in the store is instrumented with a sensor, e.g., an RFID tag, that can be used for product search and identification.

Since the initial cost of instrumentation and subsequent maintenance are two important factors that often prevent the adoption of assistive technologies, ShopTalk’s design makes a commitment to zero additional hardware instrumentation beyond what is already installed in a typical supermarket. This can be accomplished because ShopTalk is built on the explicit assumption that simple verbal route directions and layout descriptions can be used to leverage the everyday O&M skills of independent travelers to successfully navigate in the store.

In ShopTalk, the environment is represented in two data structures. The first data structure is a topological map of the locomotor space. ShopTalk’s topological map is given in Figure 3.2 with a general map of the store in Figure 3.3 for comparison. The topological
map is a directed graph whose nodes are decision points: the store entrance, aisle entrances, and cashier lane entrances. In the case of the stores where the experiments were performed, the decision points were locations where turns may have been needed to be executed. In a different store, other decision points could be added as needed, for example, when describing the area around a complicated obstacle. The edges in the graph are labeled with directions. Due to the regularity of modern supermarkets and the constraints of the problem (small-
scale shopping for items stocked on aisle shelves), three directional labels left, right, and forward were found to be sufficient. The topological map is the only software instrumentation requirement for ShopTalk to become operational. The map is built at installation time by walking through the store, noting decision points of interest, and then representing them in the map.

The second data structure is designed to take advantage of the inventory systems already used by many large- and medium-sized grocery stores. These inventory systems place barcodes on the shelves immediately beneath every product as seen in Figure 3.4. Shelf barcodes assist the store personnel in managing the product inventory. When the locations of all shelf barcodes in the store are known, this information can be used to guide the shopper through the haptic and search spaces to the location of the target product on the shelf because, under most circumstances, products are located directly above their corresponding shelf barcodes.

The second data structure is called the barcode connectivity matrix (BCM) and each of its records represents a shelf barcode and its location in the store. Each shelf barcode is associated with seven pieces of information (see Figure 3.5 and Table 3.1).

1. Aisle number. Each product is located in an aisle of the store.
2. Side of the aisle. An item is either on the left or the right side of its aisle. In order that the side remains consistent in the instructions given by the device, even as a person’s orientation changes, the side is labeled as if the shopper is standing at one particular end of the aisle and facing into the aisle. During the experiments described later, “left” and “right” were determined as if the shopper was standing at the aisle entrance nearest the front of the store and facing into the aisle.

3. Shelf section. A shelf section is a group of shelves approximately 1.22 meters (4 feet)
wide and includes the shelves from the top shelf to the bottom shelf. Shelf sections are numbered so that shelf section 1 is the section closest to one of the ends of the aisle. In the experiments, as with the side of the aisle, shelf sections with id 1 were the shelf sections nearest the front of the store. As one moved to the back of the store, shelf section numbers increased up to id 20, which was the shelf section at the opposite end of the aisle. All aisles happened to have the same number of shelf sections on each of their sides.

4. Shelf number. The product resides on a specific shelf in a shelf section. The top shelf in a shelf section is always labeled as shelf 1, with the shelves below increasing in number.

5. Relative position on the shelf. This position is not a 2D coordinate measured in some unit of distance unit, but a relative position based on how many products are on the same shelf. The left-most product on a shelf is located at position 1, and as one moves right, the position number increases.

6. Shelf barcode. This is the value on the inventory system’s shelf sticker. Shelf barcodes are used to uniquely identify each product in both the store inventory system and in ShopTalk.

7. A brief text description of the product. This description is added to the scanning instructions given to the user as shelf barcodes are scanned and products are located. In the experiments described here, the description was the name of the product.

The BCM could conceivably be built automatically from the store’s inventory database, assuming the store inventory system has all the necessary location information. However, we were not granted access to the inventory management system of Lee’s Marketplace, because it was the supermarket’s policy not to give access to their pricing information to third parties. The management described, in generic terms, what information their inventory database contained and allowed us to scan all shelf barcodes in three (aisles 9, 10, and 11) of the twelve grocery aisles. A simple graphical user interface (GUI) (see Figure 3.6) was
developed for entering the necessary information associated with a specific barcode and proceeded to scan all items in three aisles. It took 40 hours to scan and manually enter all the information for the 4,297 products found in the three aisles.

It should be noted again that the GUI development and the manual entry of product information was done exclusively for research purposes because we were not granted permission to the store’s inventory management system. A production version of ShopTalk is envisioned not as a stand-alone product, but rather as an integrated component of a store’s inventory system. The manual efforts needed in this work to build and maintain the BCM would be unnecessary in a system owned by the store. Instead of building the BCM manually, it would computed automatically from the contents of the store’s inventory database. Integration would enable updates in the inventory system’s database to be propagated to ShopTalk automatically. Employees would not be required to update the system automatically. As a component of the store system, ShopTalk would also have access to other information in the database such as pricing information.

3.3.1 Hardware

ShopTalk’s hardware (see Figure 3.7) consists of several off-the-shelf components. The computational unit is an OQO model 01, which is an Ultra-Mobile PC running Windows XP. The unit is 12.4 cm x 8.6 cm x 2.3 cm, which allows it to be carried easily. A Belkin 19-key numeric keypad is used for user input. Its #5 key has a small bump which helps the user
orient his hand to the keypad’s layout. The barcode scanner is a Hand Held Products IT4600 SR wireless barcode scanner and its base station. The scanner communicates wirelessly with the base station, and the base station connects to a computer through a USB connection. A USB 4-port hub connects the keypad and the scanner’s base station to the OQO. Since the system gives speech-based instructions to the user, the user wears a small headphone attached to the OQO.

To help carry the equipment, the user wears a small backpack. The numeric keypad is attached by velcro to one of the backpack’s shoulder straps. To ensure adequate air circulation, the OQO is placed in a wire frame attached to the outside of the backpack’s rear side. The remaining components - the barcode scanner’s base station, the USB hub, and all connecting cables - are placed inside the backpack’s pouch. When the system is in use, the user will typically place the barcode scanner in the shopping basket.

The barcode scanner (see Figure 3.8) had been slightly modified in experiments for RoboCart [37]. The modification takes advantage of the fact that many supermarkets have
shelves that curl down and have a small lip at the bottom. Plastic stabilizers, attached
to the front of the scanner, allow the scanner to rest on the shelf lips when a shopper is
scanning barcodes. The stabilizers make it easier for the shopper to align the scanner with
shelf barcodes thereby reducing the time to achieve a successful scan.

The components listed here are but one possible hardware realization of a ShopTalk-
based shopping aid. The present realization was a direct consequence of budgetary con-
straints and equipment at hand. The hardware could be further miniaturized and the
hardware costs reduced. Instead of using an OQO as the computational unit and a sepa-
rate keypad as a user input device, the software could be installed on a mobile smartphone.
The barcode scanner could be replaced with a smaller model which supports Bluetooth and
does not require a base station. Finding an optimal hardware configuration is an ergonomic
problem that is being addressed in ongoing research. The primary objective in this investi-
gation was to test the validity of the hypothesis that verbal route descriptions and barcode
scans are sufficient for independent shopping for shelved items in a supermarket.

3.3.2 Verbal Route Directions

When guiding the shopper in the locomotor space, ShopTalk issues route instructions
in two modes: location unaware mode (LUM) and location aware mode (LAM). The LUM is used in the locomotor space, whereas the LAM is reserved for the search and haptic spaces. The LUM instructions are generated from the topological map of the store and a database of parametrized route directions. A parametrized route direction is an expression such as *Turn X* or *You are at X* where X can be replaced with a context-sensitive word or phrase.

Given the start and end nodes of a path, the actual route directions are constructed from the database of parametrized route directions by replacing the parameters with appropriate fillers from the topological map. The following illustrates how a shopper interacts with the LUM instructions. These instructions would be generated if the shopper was in the middle of aisle 9 and needed to be guided to the entrance of aisle 11.

1. Shopper presses the keypad’s *Enter* key to request the next instruction.

2. ShopTalk gives the instruction: “Turn right. Walk forward to the entrance of the aisle.”

3. Shopper reaches the aisle entrance and presses *Enter* again.

4. ShopTalk gives the instruction: “Turn right. You are at entrance to aisle 9. Walk forward until you detect the entrance to aisle 11.”

5. Shopper walks, counting aisle entrances, until he reaches the entrance to aisle 11.

This mode is location-unaware because the system itself is unaware of the shopper’s actual location and orientation. Since there are no external sensors for detecting the shopper’s current location, ShopTalk explicitly relies on the shopper’s orientation and mobility skills. It assumes that the shopper can detect the environmental cues needed to make sense of the provided verbal instructions and localize herself in the environment. For example, in the above example, the first instruction assumes that the shopper is facing the shelf after retrieving a product and putting it into his basket. If the shopper is already facing the entrance of the aisle when the instruction is given, the first turn instruction can be ignored by the shopper. As one participant put it to us in an informal conversation, “I use only
those instructions that make sense to me at the moment and ignore the rest." That is the point of ShopTalk: it is always the shopper who makes the final decision about what action to execute.

ShopTalk’s LUM is conceptually no different from being guided on a cell phone by a fellow shopper who knows the store. There is some research evidence [34, 66] that people with visual impairments share route descriptions and guide each other over cell phones. The most intriguing aspect of this information sharing is that the guide is not present en route and the guidee must make sense of the guide’s instructions independently. Prior to testing ShopTalk, it was hypothesized that, over time, as the shopper uses the system repeatedly in the same store, the shopper would internalize the store layout and would not need the LUM instructions in the locomotor space at all. Therefore, the LUM instructions are given to the shopper only when the shopper presses the Enter key. If the routes and the store layout have been internalized, the shopper can choose not to request the LUM instructions. As shown in the Results section (Section 3.4.2), participants did take advantage of this feature as they learned the store layout and over multiple runs requested fewer LUM instructions.

The LUM guidance may lead to occasional navigation problems. The shopper may miscount waypoints, be blocked by obstacles such as displays or other shoppers, or have to deal with unforeseen events such as a large cart blocking the aisle while the staff is restocking inventory. Any of these situations could result in temporary disorientation. Of course, travelers with visual impairments routinely face these problems while interpreting someone else’s verbal route directions on the street. The key difference is that with ShopTalk the shopper can immediately determine his present location so long as a barcode can be scanned.

A barcode scan switches the mode of instruction from the LUM to the LAM. As soon as a barcode is scanned, the exact location of the shopper is known to the system and the system issues location-aware instructions on how to get to the target product. If the shopper becomes lost in the locomotor space, a barcode scan will inform him about his present location. Thus, barcode scans function not just as product identification cues in the
search and haptic spaces but also as error correction cues in the locomotor space. In the search space, barcode scans index into the BCM enabling the system to gradually guide the shopper through the search space to the haptic space. In the search space, the shopper is guided to the correct shelf section with instructions such as “Move dir num shelf sections” where dir is “left” or “right” and num is the number of shelf sections to move. These parameters are determined based on the target product’s barcode and the barcode just scanned by the shopper. Once the shopper is standing in front of the correct shelf, the target product is in the shopper’s haptic space. The process of scanning barcodes guides the shopper to the exact shelf with the product and finally the target product’s location on the shelf.

When the last product on the shopping list is located, the system gives the shopper verbal instructions describing how to get to the cashier. This location is somewhat easier to find than product locations due to the fact that there are specific types of noises associated with the cashier area that assist the shopper in finding it. There are also cashiers working who can assist the shopper in locating an aisle that is open.

It is unlikely that ShopTalk will be of use to all people with visual impairments. The system is designed for mobile individuals who have sufficient O&M skills to navigate indoor environments independently and have no other impairments that could potentially impede navigation such as serious cognitive or physical disabilities. ShopTalk is not designed to address physical limitations that may prevent individuals from grocery shopping, such as not being able to reach high or low shelves or being unable to carry heavy items. The level of visual impairment is not a factor in that ShopTalk can be used by people with complete or partial vision loss. ShopTalk also does not place any limitation on any additional navigation assistance the user currently uses. If a shopper uses a white cane or a guide dog, they can continue to do so when using ShopTalk.

### 3.4 Experiments

In order to evaluate ShopTalk’s feasibility and effectiveness, two experiments were performed. The first was a single-participant pilot study. The second experiment was
a multiple participant study. In both studies, ShopTalk was evaluated as an aid for a small-scale shopping task. Both experiments were performed in Lee’s Supermarket, a local supermarket in Logan, Utah.

3.4.1 Single-Participant Pilot Study

ShopTalk was first tested [97] within a single-participant pilot study. The main purpose of this study was to determine if a system such as ShopTalk was feasible. The study was designed to test three hypotheses:

- **Hypothesis 1 ($H_{S1}$):** A blind shopper with independent O&M skills can successfully navigate the supermarket using only verbal directions.

- **Hypothesis 2 ($H_{S2}$):** Verbal instructions based on run-time barcode scans are sufficient for target product localization in an aisle.

- **Hypothesis 3 ($H_{S3}$):** As the shopper repeatedly performs the shopping task, the total traveled distance approaches an asymptote.

To test these hypotheses, the BCM was built for one aisle, aisle 9, in Lee’s Marketplace. 1,655 shelf labels were scanned and the BCM field data was entered for each item. Since the BCM was built manually, the majority of the product names were left empty in order to speed up the process of entering data. However, the names of 297 products were entered, approximately one product from every shelf in each shelf section.

Several product sets were selected from the set of scanned products with names. A product set was a set of three randomly chosen products from the set of 297 products which had had their names entered during the BCM data entry. Each product set had one item randomly chosen from the aisle’s front, middle, and back. Three product sets contained items only from the aisle’s left side, three sets contained items only from the aisle’s right side, and one contained two items from the left side and one from the right. To make the shopping task realistic, each product set contained one product from the top shelf, one product from the bottom shelf, and one product from a middle shelf.
The participant was an independent blind (only light perception) guide dog handler in his mid-twenties. In a 10-minute training session before the first run, the basic concepts underlying ShopTalk were explained to him to his satisfaction. A run consisted of the participant starting at the entrance of the store, traveling to the target aisle, aisle 9, locating the three products in the current product set, and, after retrieving the last product in the set, traveling to a designated cashier lane. Figure 3.9 shows the approximate route taken during all runs with sample locations for three of products from one product set.

**Results**

Sixteen runs were completed with at least one run for each product set in five one-hour sessions in a supermarket (see Table 3.2). Ideally, the same number of runs for each product set would have been performed. However, because the experiments were performed during normal business hours with both customers and employees present, the store requested that the experiments begin after 10 PM. Because of this limitation, some runs were not repeated the desired number of times.

All three of the hypotheses appear to be reasonable for this participant. First, the participant was able to navigate to the target aisle and each target space using ShopTalk’s
Table 3.2. Number of Completed Runs for Each Product Set.

<table>
<thead>
<tr>
<th>Product Set</th>
<th>Product Location</th>
<th>Completed Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Left Side</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>Left Side</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Left Side</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Right Side</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Right Side</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Right Side</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Both Sides</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 3.10. Distances walked for each product set.

verbal route directions. Second, using only ShopTalk’s search instructions based on the barcode map and runtime barcode scans made by the participant, he was able to find all products for all 16 runs. Third, the participant’s overall navigation distance decreased with each subsequent run for the same product set. Of course, given that this was a single participant study and there were a limited number of runs, one cannot make too many generalizations based on this study. However, the success within the limited scope led to the multi-participant study reported in the next section, which was able to suggest wider support for the approach.

Figures 3.10 and 3.11 both show the downward trend in distance. Figure 3.11 also shows
the downward trend in time. The first run took the longest, 843 seconds, and had the largest distance, 376 feet. But after the second run, all times were less than 460 seconds, and all distances were less than 325 feet. The two exceptions in terms of distance were runs 7 and 13. In both these runs, the participant initially entered an incorrect aisle. After scanning a product in the incorrect aisle, the participant was instructed he was in the wrong aisle and given route directions to the correct aisle. Although, the distance increased dramatically in these runs, the time did not. The suspected reason for the lack of increase in time is that at this point the user had enough spatial knowledge of the store and understanding of the shopping task with ShopTalk that he was walking faster and searching for items faster than during the initial two runs.

Product set 5 involved walking the longest distance of all the products sets. When the same route was walked by a sighted person, the distance was 298 feet. The shortest run for product set 5 was 313 feet (see Figure 3.12). So once this participant is familiar with the environment, it appears it may be possible for him to achieve walking distances that are comparable, albeit slightly longer, to those of a sighted person.

Although the user was twice able to find a product on the first scan, on average it took 4.2 barcode scans to find the target product. Figure 3.13 shows an example of the
search the user performed for a product. The figure shows how the participant was guided to the specific location of the target product. Obviously, there is no comparison for this to a sighted shopper, because the barcode scanner is unnecessary. Therefore, in an ideal situation, the shopper with a barcode scanner would only need one scan per item.

Figure 3.13 also helps to illustrate the user moving through the three spaces. Initially,
Table 3.3. Demographics of Participants in the Multiple Participant Study.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age</th>
<th>Vision Level</th>
<th>Usual Aid</th>
<th>Aid During Experiment</th>
<th>Other Disability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>female</td>
<td>31</td>
<td>low</td>
<td>cane</td>
<td>cane</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>male</td>
<td>24</td>
<td>none</td>
<td>dog</td>
<td>dog</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>female</td>
<td>20</td>
<td>none</td>
<td>both</td>
<td>cane</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>18</td>
<td>low</td>
<td>cane</td>
<td>cane</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>male</td>
<td>29</td>
<td>low</td>
<td>dog</td>
<td>dog</td>
<td>no</td>
</tr>
<tr>
<td>6</td>
<td>male</td>
<td>23</td>
<td>none</td>
<td>both</td>
<td>cane</td>
<td>no</td>
</tr>
<tr>
<td>7</td>
<td>male</td>
<td>24</td>
<td>low</td>
<td>cane</td>
<td>cane</td>
<td>no</td>
</tr>
<tr>
<td>8</td>
<td>female</td>
<td>26</td>
<td>low</td>
<td>cane</td>
<td>cane</td>
<td>no</td>
</tr>
<tr>
<td>9</td>
<td>male</td>
<td>22</td>
<td>low</td>
<td>dog</td>
<td>dog</td>
<td>no</td>
</tr>
<tr>
<td>10</td>
<td>female</td>
<td>18</td>
<td>low</td>
<td>none</td>
<td>none</td>
<td>no</td>
</tr>
</tbody>
</table>

the user is told that the target product is located in aisle 9 on the left side in shelf section 7. Moving through the locomotor space the participant navigated to the general area where he thought shelf section 7 was located. Upon scanning a product in shelf section 6, product 1 in the figure, he had moved into the search space. He was near the target product’s shelf section but still needed to perform a small amount of locomotion in order to move so that he was positioned in front of the correct shelf section. After being informed by the system to move one shelf section to his right and then performing the action, he scanned another product, product 2 in the figure. At this point, he was now in the haptic space and no further locomotion was needed. He now was instructed to scan the lower shelf and was then guided to the product’s correct location on the shelf.

3.4.2 Multiple Participant Study

After completing the single-participant study, a similar study involving 10 participants with a visual impairment was conducted. The participants were recruited by referral from the greater Logan area of Utah with the help of the Logan, Utah Chapter of the National Federation of the Blind. All participants had received O&M training sometime in the past. Table 3.3 shows the collected demographic data on the participants. Each participant was
Table 3.4. Number of Items Scanned for the Experiment.

<table>
<thead>
<tr>
<th>Aisle</th>
<th>Number of Barcodes</th>
<th>Names Recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1,569</td>
<td>197</td>
</tr>
<tr>
<td>10</td>
<td>1,073</td>
<td>200</td>
</tr>
<tr>
<td>11</td>
<td>1,655</td>
<td>297</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,297</strong></td>
<td><strong>694</strong></td>
</tr>
</tbody>
</table>

paid a $45 honorarium for their time.

In this study, the locations of shelf barcodes were recorded for three aisles (9, 10, and 11) in Lee’s Marketplace. Table 3.4 shows the number of items that were scanned in each of the three aisles in order to build the BCM. While scanning the barcodes, the names of 694 products were recorded. From these named products, one product was randomly chosen from each aisle for a total of three products. These three products represented the product set for this experiment. Unlike the single participant study, there was only one product set for this study.

As is often the case with studies involving participants with visual impairments, it is not feasible to test, in a statistically significant way, all contributing factors in a single study due to the uneven distribution of the population in the U.S. of those with visual impairments, with the majority living in just a few urban areas. Therefore, the hypotheses below address only a fraction of the factors outlined in the previous sections.

- **Hypothesis 1** ($H_{M1}$): Using only verbal route directions, a person with visual impairments can successfully navigate the locomotor space in a grocery store.

- **Hypotheses 2** ($H_{M2}$): Verbal instructions based on barcode scans and the BCM are sufficient to guide shoppers with visual impairments to target products in the search and haptic spaces.

- **Hypothesis 3** ($H_{M3}$): As participants repeatedly perform a shopping task, the total distance they travel approaches the distance traveled by a blind shopper being guided by a sighted person.
• **Hypothesis 4** ($H_{M4}$): As participants repeatedly perform a shopping task, the total time taken to find products approaches the time needed by a blind shopper being guided by a sighted person.

• **Hypothesis 5** ($H_{M5}$): As participants repeatedly perform a shopping task, the number of barcode scans needed to find a target product decreases.

The experiment was performed during the grocery store’s normal business hours. To minimize impact on the store staff and customers, experiments began at 9:00 PM and ended between 10:30 PM and 11:30 PM, depending on the participant’s performance. Participants were given a one hour training session during which the system, the guidance techniques used, and the basic store layout were explained to them.

After the training session, each participant was led to the front of the store near its entrance and was given a shopping basket to carry. The participant was then asked to perform five runs of the experiment’s shopping route. The route (see Figure 3.3) began at the entrance of the store, went to each of the three products, and ended at the entrance of the cashier lane. Participants were not informed before starting the runs for which products they were going to shop.

During each run, participants were accompanied by two assistants. The first assistant monitored the participant’s safety and recorded observations. The second assistant followed the participant with a Lufkin measuring wheel to measure the exact distance walked by the participant. The participant was asked to press Enter on the numeric keypad at the following times: when starting a run, when ready for the next locomotor instruction, after placing a found product in the shopping basket, and when reaching the entrance to the cashier lane. The system also recorded the time and barcode number whenever the participant scanned any barcode.

When a participant found the correct barcode for a product, the participant would pick up the product immediately above the barcode and place it in the shopping basket. This run was repeated five times for each participant. After each run, the participant would
return to the store entrance to start the next run. All participants shopped for the same products in the same order and the same number of times.

Results

Repeated measures analysis of variance (ANOVA) models were fitted to the data using the SAS system. Independent variables were gender, age, vision level (low or none), O&M training (yes or no), usual navigation aid (cane, dog, or none), aid in experiment (cane, dog, or none), other navigation disabilities (yes or no), education level, and self-reported navigation skill level. Replication was achieved with ten participants, and each participant made five runs, providing a repeated measures factor "runs" with five levels as the within-subjects factor. When participants reported their vision level, if they reported any level of vision they were recorded as low; only those reporting complete vision loss were recorded as none. The dependent variables were time and distance. The time variable included the time taken by the participant to reach the required aisles and the product search time taken to find the target products.

The overall success rate of product retrieval was 100%. All ten participants were able to find all three products in every run. Verbal route instructions and barcode scans appeared to be sufficient to navigate the store and retrieve target products in grocery aisles. Thus, the null hypotheses associated with research hypotheses $H_{M1}$ and $H_{M2}$ were rejected for the sample, and experimental evidence indicates that both $H_{M1}$ and $H_{M2}$ hold.

To test hypotheses $H_{M3}$ and $H_{M4}$, a baseline run was obtained by having a sighted person guiding a completely blind guide dog handler on a shopping run. The guide knew the store well and led the shopper to the same products in the same order as was done during the experiment with the other participants. During the baseline run, the participant’s guide dog followed the sighted guide. When a product was reached, the sighted guide would place the product in the basket the participant was carrying. The baseline run was performed once, took 133 seconds to complete, and had a distance of 117 meters (384 feet).

Figure 3.14 shows that the average total time for run completion exhibited a downward trend over repeated runs. The decrease in run time across all five runs was found to be
significant with $F(4, 9) = 31.79, p < 0.0001$. Pairwise run time mean differences show that run 1 took significantly longer than the other runs. Run 2 took significantly longer than runs 4 and 5, but not run 3. Run 3 took significantly longer than run 5 but not run 4. Runs 4 and 5 effectively had the same average time.

The participants’ average run time, averaged over all five runs, differ significantly among the ten participants with $F(9, 36) = 22.62, p < 0.0001$. This is not surprising, given that the participants ranged from those with complete blindness who required a cane or guide dog to one participant who had enough vision to navigate without an aid. Thus, there appears to be sufficient evidence in the data to reject the null hypothesis associated with research hypotheses $H_{M3}$ and $H_{MA}$.

Further analysis was performed using repeated measures ANOVA to test several post hoc hypotheses comparing the total run time against several of the demographic factors. The model fitted in every case was a repeated measures model between subjects factors selected from the list above: vision level, gender, age, and so on. Each of the ten participants performed the run five times, thus there are 50 observations and 50 total degrees of freedom in the analysis. No significant effect was seen when looking at gender, age, O&M training, primary navigational aid, education level, or the self-reported navigational skill level. Level
of blindness did have an effect on the run time, which was expected. When the total run time was averaged over all five runs, the average time for the completely blind, 590.5 seconds, and the low level, 280.8 seconds, differed significantly with $F(1, 8) = 21.45$, $p = 0.0017$. When the run number was considered, the effect of level of blindness is still significant but less so with $F(1, 9) = 65.83$, $p < 0.0001$. The reason for the lower amount of significance was due to the longer times taken by the completely blind participants on runs 1 and 2. It appears that over time the completely blind participants were able to increase their efficiency and began to approach the performance levels of the low vision participants.

Like the average total time, the average distance traveled during a run fell as the runs were repeated (see Figure 3.15). The analysis of the average distance is analogous to the analysis of the average total time. Effects of the run number and level of blindness were significant, with $F(4, 9) = 5.52$, $P = 0.0159$ and $F(1, 9) = 53.13$, $p < 0.0001$, respectively. The total distance decreased significantly with run number, and the total distance was significantly greater for the completely blind participants. This analysis indicates that, as the participants gained experience, their routes contained smaller errors related to distance. As discussed later in Qualitative Observations, it appeared that some of the increased accuracy came from learning the location of landmarks, which helped the participants to
Figure 3.16. Average number of items scanned over repeated runs. The baseline is not shown since products were located visually, not scanned, by the sighted guide in the baseline run. The ideal would be three products scanned.

make better distance judgments.

Figure 3.16 shows that the average number of products scanned per run also fell over repeated runs. The decrease in the number of products scanned across all five runs was found to be significant with $\chi^2(4) = 24.26, p < 0.0001$ allowing the null hypothesis associated with $H_{M5}$ to be rejected. This indicates that as the shoppers gains more experience, they became more efficient with the scanning process. The number of products scanned in the first two runs, when averaged over all ten participants, did not differ significantly, but the average number of products scanned on runs 3, 4, and 5 were all significantly lower than run 1, $\chi^2(1) = 7.28, P = 0.0070, \chi^2(1) = 13.23, P = 0.0003, \chi^2(1) = 16.12, p < 0.0001$, suggesting that by runs 3 or 4 the participants were approaching the asymptotic limit for the number of products they would need to scan on this particular route.

The level of blindness, complete or low vision, caused a significant effect on the number of items scanned across all five runs, $\chi^2(1) = 9.63, P = 0.0019$, with low vision shoppers scanning fewer numbers of items. The implication is that because partially sighted shoppers have additional sensory information not available to the completely blind shoppers, they are more efficient in their barcode scanning. Depending on the amount of partial vision
and the type of visual impairment, some shoppers may not have to always use the barcode scanner to find products. The difference in the rate at which the two groups improved over the five runs was not found to be significant.

The times when ShopTalk provides LUM instructions can be divided into two different groups. The first group of LUM instructions are given whenever participants place a found target item into their basket. In order to signal that they are ready for the next LUM instruction that will guide them to either the next target product or the cashier, they pressed the keypad’s Enter key to instruct the system to give the next instruction. These instructions are mandatory. For all users, there will be at least one mandatory LUM instruction for each target product and one for the cashier.

The remaining group of LUM instructions are optional. The user can choose whether or not they want to hear these instructions. To request one of these instructions, the user presses the keypad’s Enter key whenever they reach a decision point indicated by a previous LUM instruction. For example, one type of instruction indicates how to get from the entrance of one aisle to the entrance of another aisle. When users reach the target entrance, they have the option of pressing the Enter key to receive the next set of LUM instructions that will guide them into the aisle. The hypothesis was that as users gained experience with the store layout, they would not need to access the optional LUM instructions and would begin using his mental map of the store. During the experiments, the number of times each user requested optional LUM instructions was recorded. Figure 3.17 shows that the average number of LUM instructions fell over repeated runs.

**Workload Analysis with NASA TLX**

After completing the five runs, participants were asked to complete the NASA TLX (Task Load Index) [44] questionnaire, which allowed us to analyze the perceived workload of the shopping task with ShopTalk. The questionnaire asks participants to first score six dimensions on a scale of 1 to 20. The dimensions as described in [44] are:
Figure 3.17. Average number of optional LUM instruction requests. The baseline is not shown since a sighted guide does not require verbal instructions. The ideal would be zero requests.

- Mental Demand: level of mental activity required of the participant during the task (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)

- Physical Demand: level of physical activity required during the task (e.g., pushing, pulling, turning, controlling, activating, etc.)

- Temporal Demand: the amount of time pressure felt during the task

- Performance: how successful did the participant feel in achieving the goals of the task

- Effort: how hard the participant worked, mentally and physically, to complete the task

- Frustration: what level of discouragement versus satisfaction the participant felt in completing the task

Once the six dimensions were rated, participants were asked to choose the most important contributor to the overall workload. During this part of the questionnaire, each dimension is randomly paired with the other five dimensions and the participant selects one
of the two choices. In this manner, all dimension pairs are presented and scored by the participant.

The overall task workload score is computed from all pairwise combinations. The six ratings are first scaled from the range \([1, 20]\) to the range \([1, 100]\) to give the dimension’s value. Next, the number of times a dimension is selected as the most important workload contributor in the pairwise questions is counted. This count for each dimension is scaled to the range \([0, 0.333]\). A 0 signifies that the dimension was never a major contributor, i.e., the dimension was not chosen in any of its five comparisons. A .333 signifies that the dimension was chosen as the major contributor in all five of its pairwise comparisons. This scaled count is the dimension’s weight. The final workload score is calculated by multiplying each dimension’s value by its weight and then adding together the resulting six products. Table 3.5 shows the averaged weights, values, and final workload score for the set of participants.

The TLX workload score was found to be significant with respect to the total run time averaged over all runs for all ten participants, \(F(1, 9) = 17.09, p < 0.0025\). There was also a significant interaction with the run number with \(F(4, 9) = 8.05, p < 0.0048\). On each run, a higher score corresponds to a higher average run time, all significant at \(\alpha = 0.05\). Runs 3 and 5 were significant at \(\alpha = 0.01\). In general, users who perceived this task as difficult took longer to complete the route.

**Qualitative Observations**

In addition to the statistical analysis reported, a number of qualitative observations were made regarding the participants, the way they used the device, and how they interacted with the store environment. This section discusses these observations and some of their implications.

**Shopping Techniques.** All participants were instructed on how to use ShopTalk, yet several participants were observed using individual methods to enhance their search techniques. Two participants used touch to help locate their position relative to a product. Participant 1 noticed that marshmallows were located next to product 1, and by the fourth
run was locating the marshmallows with her hand and then moving from that position to the target product. Participant 3 judged distances on her first run by running her finger along the shelf edges. There is a separator between shelf sections which she counted in order to determine her position in the aisle.

Two other participants developed a different technique for finding product 1. Product 1 was near the end of the first aisle, towards the back of the store in shelf section 18. Instead of beginning their search at shelf section 1, they would walk to the opposite end of the aisle, shelf section 20, and then would backtrack to find shelf section 18. For these participants, determining the short distance from the back entrance of the aisle was easier than determining the long distance from the front entrance to the aisle.

Although participant 8 reported only being able to see light and dark, she was able to find product 1 on her first try with little apparent difficulty. During training, participants were told that a shelf section was approximately 1.22 meters wide. She reported taking advantage of this knowledge by counting her steps, allowing her to achieve a good level of accuracy without using the touch techniques used by others. She did use touch to help locate product 2, where she noticed that the shelf section next to product two was covered by a plastic divider which was different from the usual flat, metal shelves used in the majority of the store.

In each of these cases, the shoppers devised their own methods for accomplishing the task; none of these techniques were mentioned during training. ShopTalk provided them
with enough flexibility that they could enhance their performance with a technique that worked for their navigation skill level and ability. One advantage of tracking these techniques is that over time, the techniques which are seen to be the most useful could be explained to new users of the system, increasing the effectiveness of ShopTalk for all users.

**Environment.** Because the reported experiment was performed in a real-world supermarket, the store environment led to several noteworthy occurrences. In general, store employees and other shoppers would move out of a participant’s way if they noticed the participant. At times, though, other people did not notice the participant and the participant would have to adjust. For example, an elderly man walking slowly down an aisle slightly slowed the walking speed of one participant. Ultimately, none of the situations in which another person could be considered an obstacle prevented a participant from accomplishing the task.

Another issue was that late evenings, when the experiments were performed, were also the time when employees restocked shelves. The employees would start placing large boxes in aisles and wheeling large carts with boxes around the store and into the aisles. The carts would stay in one spot for several minutes as the employees placed items on the shelves. If an employee noticed that a participant needed to pass the employee or to enter an aisle that a cart was blocking, the employee would move the cart and let the shopper pass. In several cases, however, the employee did not notice a participant. One participant, when repeatedly blocked from entering an aisle, went to the next aisle, proceeded to the back of the store, and returned to the desired aisle from the other end. Thus, the participant was using her own mental map and expectations of the store layout instead of following ShopTalk’s instructions. In fact, ShopTalk’s instructions never referred to the back of the store. All instructions regarding entering and exiting aisles were given solely in relation to the front of the store and entrances to the aisles closest to the front of the store.

To show items on sale, Lee’s Marketplace sometimes stacks them in the aisles (see Figure 3.18). The stacks and other promotional displays proved to be both advantageous and problematic. The third product in the experiment happened to be located two shelf
sections after a large stack of cans of spaghetti sauce. Several of the participants were observed searching for this stack to help them locate the last item. The use of these stacks as landmarks is questionable over the long term since the items for sale change over time, some stacks gradually disappear and new ones appear at different locations. Stacks and other product displays also have to be negotiated carefully. Two participants accidentally hit the stack of spaghetti sauce cans with each knocking a can of sauce to the floor. Had the products been glass jars rather than metal cans, the result may have been different when the product hit the ground, potentially embarrassing the participant. The end of the aisles where large displays of sales items are placed also provided similar challenges due to the differences of height of different displays. Participants would occasionally nudge items in these display areas as they walked by. Although the improvement at detecting stacks was not quantified, it appeared that over repeated runs the participants became better at avoiding product stacks as they learned their locations.

**Ergonomics.** The experiment was designed to determine the feasibility of ShopTalk. However, the experiment also revealed some ergonomic issues with the device. The main issue mentioned by four participants during informal discussions was there was a lot of equipment to manage. In addition to their usual cane or guide dog, the participant had
to manage a barcode reader, a shopping basket, and press buttons on the keypad. One participant noted that in a real shopping situation, she would also have to manage her preschool children. As noted previously, miniaturizing the hardware is a research priority.

The barcode scanner was typically, but not always, carried in the basket. When scanning for items, the guide dog handlers would instruct their guide dog to sit. Cane users would lay their cane on the ground, rest their cane on the shelves, or change their grip in order to scan for items. Several participants, however, seemed to be unaware of how their canes, dogs, and shopping baskets were extending beyond their personal space. For example, participant 4 held his cane and the shopping basket while he scanned barcodes on the shelves. Because he was concentrating on the scanning process, he did not seem to be aware of the fact that his cane and basket were hitting items on the shelf. While no items were knocked off the shelves, it is possible that an item could be knocked off a shelf. Another participant who used a guide dog would place his dog between himself and the shelves in order to keep the dog under control. It is possible that other shoppers may not appreciate the fact that the dog was touching items on the shelf as it sat there. Two participants seemed unaware that products were stacked on one another. In these cases, they would pull out the bottom product although it would have been safer to take the top product. These types of issues would need to be mentioned in training with the system in order to make people aware of these issues.
CHAPTER 4
CRISS FRAMEWORK

4.1 Introduction

After the ShopTalk experiments, the question arose as to how far the verbal route
description only approach could be taken when guiding an independent traveler with a
visual impairment. Can it only work indoors, or could it work outdoors as well? Is there
a limit to the size and structure of an environment for this approach in order for it to be
feasible? What qualifies as a good route description for a visually-impaired traveler? How
are the route directions given by a partially-sighted individual different from those given
by a person with complete vision loss? How are the directions similar? When following
route directions, does it make a difference whether or not the person who created the route
description had the same amount and type of vision loss as the person who is traveling the
actual route?

Some of these questions have already been addressed through previous work, showing
that the route directions generated by people with visual impairments are different than
those generated by people with no visual impairment. For example, research [22] focusing
on a comparison of route descriptions from sighted children and from visually-impaired chil-
dren found differences between the two populations. When describing a route from memory,
the route descriptions from the visually impaired children were longer than those from the
sighted children. The visually impaired children also referred to tactile information in the
environment and mentioned hazards such as the top of stairs or protrusions whereas the
sighted children did not mention either of these features. In another set of studies [51, 52],
blind children were observed breaking up route directions into segments. The directions pro-
duced by the children not only mentioned landmarks at turns, but also mentioned a series of
landmarks experienced along non turning segments of the route. Route directions produced by sighted children included fewer mentions of landmarks and only when mentioning a place or a turn. In a similar study comparing visually-impaired and sighted adults [7], the route descriptions of the visually-impaired adults contained significantly more information on distance, directions, and landmarks than the route descriptions given by the sighted adults. The visually-impaired participants also mentioned obstacles along the routes, whereas there was only a single mention of an obstacle among the sighted participants. As the study concludes, “sighted persons give environment-oriented descriptions, whereas the blind tend to use person-oriented descriptions” [7:217].

The differences between route directions produced by a visually impaired traveler familiar with a route can be striking when compared with the directions from a sighted traveler. In order to help understand the difference, two people, one with total vision loss and another with normal vision, were asked to describe a route on USU’s campus with which they were both familiar. The route that they were asked to describe (see Figure 4.1) started at the USU’s Disability Resource Center (DRC), went to the Old Main building, and then
went indoors to the elevators in Old Main. The following directions, transcribed from a recording, were described from memory by the sighted person:

You go out the patio doors and then you head - there’s a huge water fountain. You’ll, like, turn left there and then you’ll come up to the Eccles Science and Learning Center. And then you’ll just go straight past that building. And then past that building, I think it’s the Education - wait, no, I think it’s the Geology building. You’ll take a right there. And then it will lead you, you go there until you see the Old Main building. Then you come up some steps and then you’ll just go down that hallway to the very end of, uh, the middle of it and there’s the elevator.

The following is a portion of the route directions given by the person with the visually impairment. Due to length, the entire set of route descriptions are not shown here, but can be read in their entirety in Appendix A. At this point in the route description, the traveler would have entered Old Main and gone through one set of doors and then a second set of doors, and would now be searching for the elevators.

...So once you enter into, enter through those doors, there will be carpeted surface and this is a main hallway heading east-west inside Old Main. There’s classrooms on both sides. Move across this room, or through this hallway and you will find that the hallway, this is a large hallway and to get to the rest of this hallway, the hallway narrows on the left side. So make a gradual turn, and if you’re using a guide dog, you can make a moving turn left and your dog will find the hallway. If you’re using a cane there, you could run into a wall that tapers the large hallway into the small hallway. So now that you’ve entered the small hallway you are still heading west inside Old Main. You will come upon an intersecting hallway. You will hear that the hallway runs north - south across the length of Old Main. Once you reach this intersecting hallway, you need to make a left turn. Now you’re heading south. Make approximately 5 to 10 steps,
you will most likely be able to hear a hallway on your left which harbors the elevators. Make a left turn into this small hall or corridor and the elevators will be on your left, on the north side. The elevators are facing south. There’s two elevators. The up/down buttons for the elevator are right in the middle of both of these elevators.

The difference between these two descriptions is not difficult to see. The amount of detail is far greater in the blind person’s description than in the sighted person’s description. The sighted person’s description is condensed and assumes a lot of spatial information, which is evident from the fact that the portion of the blind person’s description shown represents only part of the last sentence of the sighted person’s description: “... then you’ll just go down that hallway to the very end of, uh, the middle of it and there’s the elevator.” Although the descriptions were captured verbally and then transcribed, contain grammatical patterns not typically seen in written language, and may be longer than would be seen if originally written rather than spoken, the differences in length are still dramatic. The sighted person’s set of directions contains 110 words, while the blind person’s entire set of directions contains 1,534 words. The additional words represent pieces of information not contained in the sighted person’s directions. Details included in the blind person’s description include references to:

- Surface textures, such as “a cement structure,” “cobblestone, or a rougher surface,” “a tiled surface,” and “carpet.”
- Distances in units and number of steps, such as “Make approximately 5 to 10 steps” and “probably 15 to 20 yards.”
- Sounds, such as “... you can hear cues. People socializing.” and “... hear other people walking in a perpendicular fashion.”
- Odors, such as “... the smell of cigarette smoke...”
- Cardinal directions, such as “Now you are facing north.” and “... a sidewalk intersecting directly west, heading east - west...”
Another interesting thing to note about the sighted person’s directions is that they contain an error. She incorrectly named the building next to the Eccles Science and Learning Center (ESLC) as the Geology Building, when in actuality it is the Animal Science building. Although this may cause confusion for some travelers, for most sighted travelers the error would not be an issue. The ESLC building has a large sign with its name on the side of the building, providing an important visual clue to the traveler that he is in the correct location. Once the traveler turns and visually locates Old Main, he can visually home in on Old Main, largely ignoring the error in the route description. A blind traveler attempting to follow these directions would not have the advantage of the signage acting as an error correction or the ability to visually home-in on the distance visual landmark.

4.2 CRISS: VGI Websites for the Blind

Passini and Proulx [102] showed that people who have visual impairments prepare more for travel, make more decisions, and use more information than sighted travelers. One of the most common ways to prepare for travel is to use a map. There are many map-oriented websites on the Internet today, such as Google Maps [42], MapQuest [82], and Yahoo! Maps [125]. While these websites are useful to the majority of the population, people with visual impairments are generally unable to use the sites because much of the information on these sites is presented visually. Google Maps has addressed this problem to a degree. If a user enters the special URL http://maps.google.com/?output=html, the Google Maps website reverts to a simpler user interface that works better with screen readers (compare Figure 4.2 and Figure 4.3). Yet even the simplified interface is not able to be used by all visually impaired users since the most important piece of information on the screen, the map, is still a visual element.

While the map screen may be of limited use to a blind person, the same URL also allows a user to obtain route directions via a user interface more suitable to screen readers. Historically, the route directions produced by Google Maps and other map sites have been targeted to car drivers, but now users can get route directions for walking a particular route. However, these new walking directions are still of limited use to the blind population. As
Figure 4.2. Google Maps’ standard user interface.

Figure 4.3. Google Maps’ simplified user interface for screen readers.
Figure 4.4 shows, the walking directions are clearly not targeted to a visually-impaired traveler. A warning is given stating “Use caution - This route may be missing sidewalks or pedestrian paths,” but the segments with the missing features are not clearly identified in the text. Turns and distances are given, but there is no description given of the place where the turn is to be performed, making it more difficult for a visually-impaired navigator to determine when to turn. The route description is also missing information that could impact the safety of a visually-impaired traveler, such as whether the streets are one-way or two-way, how many street intersections need to be crossed during long distances, whether intersections contain stop signs or traffic lights, and descriptions of possible obstacles such
Figure 4.5. Walking route generated by Google Maps (A - B1) from DRC (A) to Old Main (B1 and B2). The second route (A - B2) is the route described by people.

as tree roots coming up through the sidewalk or areas of low hanging tree branches.

The map interface, as is commonly implemented on the web today, is not suited towards a basic map-related need of people with visual impairments, that is, getting descriptions of routes from one location to another. The user interface can be adjusted, as Google has done, to address screen readers and speech interfaces. However, the data, which is used to generate the text that users and their screen readers read, is insufficient. The geographic information system (GIS) data does not exist at a level that would allow applications to generate meaningful and safe route directions targeted to blind travelers. Current maps are primarily auto-based street level maps. Another limitation is that information for indoor environments, i.e., building maps, does not exist.

To demonstrate the effect of insufficient data has on the quality of walking instructions, Figure 4.5 shows the walking route generated by Google Maps for the route from the DRC
to Old Main, as described earlier in Section 4.1. The walking route given by Google Maps is .9 miles. As can be seen in the image, Google Maps is limited to using sidewalks that happen to map to streets resulting in a route that is not direct and leads to a different Old Main entrance than the one given by people. The human-generated route has no restriction and is the most direct route possible.

As with ShopTalk, the navigational knowledge and skills of experienced navigators could be leveraged in this situation. A person with a visual impairment, when familiar with a given geographic area, often has a wealth of spatial information about that area, particularly the routes along which he travels on a regular basis. Consider a student who has spent four years attending a university and navigating its campus. The student has learned and traveled multiple routes around the campus - classroom to classroom, building to building, home to campus and back. If the experienced student were able to document and share descriptions of the routes with which he is familiar and could make the descriptions available to new students who are also visually impaired, the new students would have an easier time learning the new routes around the campus.

### 4.2.1 Volunteered Geographic Information

The relatively new area of volunteered geographic information (VGI) [41] consists of user- and community-generated GIS tools, data, and websites. VGI websites and communities encourage interested volunteers, rather than GIS professionals, to provide the data used to build maps or other GIS services. VGI websites work on the basic assumption that individuals will voluntarily collect and upload small pieces of appropriate information. The website’s back-end software then assembles the individual collections of data into a single cohesive set of data. Although the data quality may be of a different quality than more traditional and professional sources such as the U.S. Geological Survey, VGI often represents unique sets of information not available elsewhere [26]. The theory is that people who live in a particular area will be more familiar with that area than professional GIS people working for a map-making organization located in another state or country. This familiarity allows those people to contribute information to VGI projects, information that is generally
unavailable to people who are not from that particular area nor intimately familiar with the data source.

Many examples of VGI websites exist on the Internet and often provide the same services as professionally created GIS websites. OpenStreetMap \[43,98\] is a VGI website and set of tools that has the goal of creating a large, free set of map data. In some ways, the data set is similar to the data used by commercial sites such as Google Maps; both present a vector representation of map data. Because it uses commercial sources for its map data, Google also offers services such as satellite images, a feature that is not available on OpenStreetMap. On the other hand, since OpenStreetMap is user-edited, features such as sidewalks and lesser known buildings may be added and annotated on their maps but not on Google’s maps. Wikimapia \[122\] is an example of a VGI site that takes advantage of users’ local GIS knowledge. Users annotate maps from Google Maps with additional information such as building names, descriptions, and photos. VGI websites are not limited to displays resembling traditional maps. Trailpeak \[115\], for example, allows users to add, edit, view, and download information related to trails for activities such as hiking, mountain biking, and kayaking in Canada and the United States. While a small map is displayed for a trail, site members also add text as trail descriptions, directions to the trail, reviews of the trail, and GPS waypoints.

One of the characteristics of these sites is that the users are familiar with the local data. OpenStreetMap users are familiar with the area of the city that they are mapping since they generally live or travel in that area. Users on Trailpeak are familiar with the trails they upload and edit because they actually use those trails. Because of their local experience and domain knowledge, these users can create maps and data sources that can be as credible as more traditional sources \[26\].

Independent travelers who have visual impairments represent credible sources of route knowledge for the areas where they live, work, and explore on a daily basis. Their expertise is a combination of familiarity with well-traveled routes and an understanding of the skills people with visual impairments use navigate the world and how these skills can be applied
to following known routes. If this knowledge could be shared, it may be possible for people unfamiliar with the area to use this knowledge for traveling unfamiliar routes.

We propose a framework for VGI websites meant to capture and share the local expertise of independent blind travelers. Termed the Community Route Information Sharing System (CRISS), it is intended to provide more detailed and user-appropriate levels of information to enable independent blind navigators to follow routes in unfamiliar indoor and outdoor environments without the need for external sensors such as those based around GPS, RFID, infrared, etc. Since intended CRISS users themselves would have visual impairments, would be familiar with the geographic area, and would have traveled the routes themselves many times, they could, in turn, describe routes at a level of detail appropriate for other blind navigators. The conjecture is that if a route is described with a sufficient and appropriate amount of detail, a visually impaired person who is unfamiliar with the route can use his everyday navigation skills and abilities to successfully follow the route.

A website built using the CRISS framework would allow communities of independent blind navigators to come together and work collaboratively, learning and sharing route descriptions for large geographic areas such as college campuses or even towns and cities. The spirit of these websites would be similar to community-based websites, such as Wikipedia [123], that allow users to add, create, and edit data, resulting in a large and dynamic user-managed collection of knowledge. In CRISS, the community-managed knowledge would consist of a large collection of route descriptions, pieces of natural language text that describe how to get from location to location within a predefined area. The route descriptions would cover routes that are entirely indoors, entirely outdoors, or a mixture of both. As the collection of route descriptions for a geographic area grows, the user community would end up creating a route-based map for use by independent visually impaired travelers. People new to the area, instead of relying solely on sighted guides to introduce them to the area, would be able to download user created and approved routes from CRISS. Using their everyday navigation skills and abilities, they would be able to travel the unknown route independently.
4.2.2 CRISS Data Structures

In order to model the environment as a set of route descriptions, the CRISS framework defines two main data structures. The first is a hierarchical set of landmarks that is used to define relationships between landmarks in the environment. The second data structure is a route graph that is used to represent natural language route descriptions written by CRISS users. Together, these two data structures form the core set of information that a CRISS-based website would use.

Both data structures reference landmarks. A landmark is defined as any useful location or object in the environment that a person who is visually impaired may mention in a route description and that a person traveling the route can detect. Landmarks also include both the starting locations and the ending locations of routes. Examples of possible indoor landmarks include, but are not limited to, rooms, doors, and hallway intersections. Outdoor landmark examples include entire buildings, sidewalks, and streets. Landmarks can be detected by various human senses and may therefore include entities or locations not typically mentioned by sighted travelers in their route descriptions, e.g., “when you smell bread baking” or “where you feel the carpet change to tile.”

Landmarks are given a unique id allowing CRISS to uniquely identify each landmark in the system. This also allows multiple landmarks to have the same name. Many buildings, for example, have multiple floors and may have a landmark named SECOND FLOOR. Landmarks may also have a list of zero or more alternative names, which would allow a landmark to be referred to in multiple ways. For example, the landmark ROOM 414 representing room 414 in the Old Main building could have the alternative names COMPUTER SCIENCE FRONT OFFICE, CS FRONT OFFICE, and CS DEPARTMENT HEAD’S OFFICE. Finally, a landmark has a free-text description that allows users to describe the landmark. The description allows knowledgeable users to provide information about the landmark that they feel people who are unfamiliar with the landmark may want to know.

The landmark hierarchy is a tree representing part-of relationships. Larger and more general landmarks are stored higher up in the hierarchy, and more specific landmarks are
stored lower in the hierarchy. The larger landmarks contain the more specific landmarks creating the part-of relationship. Figure 4.6 shows that as one moves down the hierarchy, the landmarks move from large areas, such as the entire university, down to landmarks marking the equivalent of a specific location, e.g., a water fountain in a specific hall on a specific floor of a specific building. The concept of hierarchical landmarks in CRISS is influenced by regions in the topological level of the spatial semantic hierarchy (SSH) [60]. Regions in the SSH are areas that can contain smaller regions or that can be part of larger regions. Landmarks in CRISS can have sets of landmarks as children or can be a child of a
larger landmark. In other words, part-of is the only relationship in the hierarchy.

Although the landmark hierarchy can be preloaded with an initial set of landmarks - building names, department names, key rooms and offices, etc. - the majority of the landmarks will be edited and maintained over time by the user community. When a new landmark is mentioned in a route description, users have the capability to add it to the landmark hierarchy if it has not been previously added. Users can also remove unnecessary landmarks if the landmark is determined not to be needed. If a real-world object or location is used in any route description or is never deemed a starting or ending location for any route, the user community is free to never include it as a landmark in the hierarchy. Users also have the capability of changing a landmark’s position in the landmark hierarchy, ensuring the landmarks are arranged in the most appropriate order.

CRISS’s second data structure used to model the environment is a set of user written natural language route descriptions. A route description is the description of how to travel from one location to another location. Users describe routes using natural language as this is a standard means that people use to communicate routes. This second data structure also allows users to describe routes in terms that are appropriate for other visually impaired travelers.

A route description has three properties: a starting location, an ending location, and a natural language description that guides a person from the starting location to the ending location. In CRISS, the starting and ending locations reference landmarks in the landmark hierarchy. The natural language description is stored as a list of route statements. A route statement is an individual sentence from the user created route description. Route descriptions are broken into route statements in order to associate landmarks mentioned in the descriptions with the normalized set of landmarks in the landmark hierarchy. Route descriptions each have a unique route id, and each route statement also has a unique route statement id.

A landmark from the landmark hierarchy associated with a route statement is considered a type of tag, or metadata. Tags are data that give meaning to other pieces of
data [83] and are in widespread use today. Flickr [27] and YouTube [126], for example, have user-created tags that describe the content of the images and videos respectively. One of the advantages of tags is that they allow users on these sites to find related content easily. For example, on Flickr, users can find all images which have the associated tag ORIGAMI.

One problem with the Flickr and the YouTube tagging systems is that they do not have a consistent naming structure and their namespaces are flat [83], resulting in little semantic meaning associated with the tags. The tag DOORS, for example, could be associated with a photo containing multiple wooden doors in a hall as well as a photo of the members of the classic rock band The Doors. One of the goals of the semantic web is to provide well-defined meaning that both computers and humans can use [5]. CRISS enforces this goal by only allowing route statements to be tagged with landmarks from the landmark hierarchy. In CRISS, a tag represents one specific landmark.

The advantage of this type of tagging is that a normalized and uniform landmark set is associated with the data, in this case, a route statement. As a result, the system ensures that multiple route statements containing phrases with different names or abbreviations for one landmark, e.g., “Computer Science,” “CS Dept,” and “cs department,” would all be tagged with a single, standard landmark tag, e.g., COMPUTER SCIENCE DEPARTMENT. Users could also tag a route statement with landmarks not explicitly mentioned in the route statement. For example, one user may add a route description with the following route statement: “Turn right when you detect the intersection with the main hall.” Another user may choose to tag this route statement with the tag WATER FOUNTAIN since that user has found that performing the action in the route statement causes them to pass by a water fountain located at that particular hall intersection.

Tagging route statements with landmarks serves several purposes. First, a traveler, when exporting a route, would optionally be able to choose to export the landmark tags along with the natural language route description. For users who have little knowledge of the area through which the route description will guide them, the extra description provided by the landmark tags will give a better idea of what to expect. Second, tags provide a uniform
way of marking landmarks in natural language statements. Multiple landmarks, even those not explicitly mentioned in the route statement, can be associated with a route. A third benefit is that algorithms can be developed to process the routes in terms of landmarks rather than in terms of natural language.

4.3 Introduction to the Route Analysis Engine

The Route Analysis Engine (RAE) has been developed to assist with two problems associated with routes in a CRISS-based system. First, route descriptions need to be broken into lists of route statements, and each resulting statement needs to be tagged with the appropriate landmarks. If the landmarks could be extracted automatically, the user’s experience when entering route descriptions would be simplified. The second problem that RAE addresses is that of finding new routes. If new routes could automatically be discovered in a set of route descriptions, it would both increase the rate at which routes are entered into the system and open up new avenues of explorations of areas.

RAE’s automated landmark tagging process is called *autotagging*. It takes advantage of information extraction (IE) techniques to process natural language route descriptions written by users. After route descriptions are broken into a list of individual route statements, each route statement is tagged with its landmarks. RAE’s autotagging process is covered in Chapter 5.

The second process is called *path inference*, which is the process of inferring a new route description from existing route descriptions by determining where existing routes intersect. For example, suppose there is a route from A to B and another route from C to D. If the two routes share a common landmark E in the middle, it may be possible to infer a new route from A to D by way of E. In addition to reducing data entry for common segments shared by multiple routes, new routes of which no user is yet aware may be discovered. RAE’s path inference process is covered in Chapter 6.
CHAPTER 5
LANDMARK AUTOTAGGING

5.1 Introduction

Written route descriptions consist of natural language text. Although this is the preferred format for humans, natural language, or unstructured free text, is not handled efficiently by computers. Computers work better when data are presented in a structured format. One set of techniques used to extract information from unstructured text into structured formats is information extraction (IE) [16].

Since IE has been researched for over a decade, it is no longer necessary to implement general IE techniques from scratch. There are now tools available that provide IE frameworks allowing researchers to concentrate on developing IE tools for specific domains. One such framework is the General Architecture for Text Engineering (GATE) system from the University of Sheffield [19]. GATE is a Java-based, general purpose natural language processing (NLP) system and includes various components for different types of NLP tasks. GATE’s IE component is A Nearly-New IE system (ANNIE). ANNIE uses the pattern matching language Java Annotations Pattern Engine (JAPE) [20], which allows an IE developer to write patterns that can extract matching information from text. JAPE is built on top of Java and is based on the Common Pattern Specification Language (CPSL) [1], a language whose purpose was to replace structures like regular expressions and system-specific formalisms with a common formalism for performing IE in many systems. A brief overview of ANNIE and JAPE is provided here. An interested reader should refer to [20] for complete details of GATE, ANNIE, and JAPE.

ANNIE consists of several processing resources (PR). Each PR is responsible for a specific IE task, and typical ANNIE usage chains the PRs together so that the output of
one PR is fed into the next PR in the chain. The default ANNIE PRs are:

1. Tokenizer
2. Gazetteer
3. Sentence splitter
4. Part-of-speech (POS) tagger
5. Named entity transducer
6. Orthomatcher

The tokenizer splits a natural language text into basic tokens such as numbers, words, and punctuation. After the tokenizer, the text is passed to the gazetteer which is used to find entities that are well-known and capable of being listed, for example, the names of all the employees in a company. After the gazetteer, the text is passed to the sentence splitter. As its name implies, the sentence splitter marks where sentences begin and end. Once the sentences are located, the text is passed to the POS tagger. The tagger assigns a part of speech to each token. Parts of speech include nouns, verbs, etc. Most parts of speech have been subdivided into more specific types for greater control. For example, a noun could be classified as a singular noun, a plural noun, a singular proper noun, or a plural proper noun.

The text is then passed to the NE transducer which is responsible for running JAPE rules that identify entities in the text. Finally, the text is passed to the orthomatcher which finds references in the text that refer to one another. For example, if a piece of text includes the strings “Barack Obama” and “President Obama,” these two strings should be marked as coreferences since they both refer to the same person.

Over time, IE techniques have been subdivided and classified into five subtasks of techniques as listed in Cunningham [18]:

- Named entity recognition (NE): Finds and classifies entities such as names, places, etc.
• Co-reference resolution (CO): Identifies identity relations between entities.

• Template element construction (TE): Adds descriptive information to NE results using CO.

• Template relation construction (TR): Finds relations between TE entities.

• Scenario template production (ST): Fits TE and TR results into specified event scenarios.

The subtask that maps to RAE’s autotagging process is named entity recognition (NE). The purpose of autotagging is to identify landmarks, whereas the purpose of NE is to identify entities. All landmarks are entities in the text, but not all entities are landmarks. An example of a non landmark entity is a person’s name. It is important not to completely disregard the non landmark entities because they can reference a landmark or be part of the landmark’s name, e.g., “John’s lab.” Entities fall into one of two groups: known entities and unknown entities. Known entities are entities that are known and do not necessarily need to be discovered through IE. For example, RAE known entities are the landmarks that have already been identified through the process of tagging other route descriptions and contained in the landmark hierarchy.

Finding known entities is straightforward and maps directly into ANNIE’s gazetteer task. A file is created listing each possible string for a given entity type. When the gazetteer PR is run, matching text is annotated with the appropriate entity type. For example, if the gazetteer list contains already contains building names and office names from the landmark hierarchy, the sentence “Dr. Kulyukin’s office is on the fourth floor of Old Main” contains references to two annotations after processing: one annotation referring to the string “Dr. Kulyukin’s office” and one annotation referring to the string “Old Main.”

Unknown entities, including unknown landmarks, are entities that are not previously known and cannot be derived from a simple look-up list. These also include words that are misspelled. If “Old Main” is in the gazetteer list but a user spells it as “Old Mian,” the user’s text will not be annotated by the gazetteer. Extracting unknown entities requires more
processing than a simple look up. RAE’s autotagging is able to achieve this by extending the JAPE rules associated with ANNIE’s NE transducer. The combination of ANNIE’s default processes and RAE-specific NE rules allows the system to find and annotate the landmarks in the natural language route instructions.

ANNIE’S NE transducer relies on rules written in the JAPE language. An interested reader is referred to [20] for complete details on the JAPE syntax. The following is a brief introduction to the syntax in order to help explain RAE’s components.

JAPE is a pattern matching language that uses regular expressions to match patterns found in natural language text. A JAPE rule consists of a left hand side (LHS) and a right hand side (RHS) separated by an arrow in the form of: $\rightarrow$. The LHS consists of patterns to look for in the text, and the RHS consists of statements that manipulate and create annotations. An annotation describes what a piece of extracted text represents, e.g., a building name or an employee’s name. The LHS of a JAPE rule uses regular expression operators to provide regular expression pattern mapping. These operators include an asterisk ( * ) for 0 or more matches, the plus sign ( + ) for 1 or more matches, a question mark ( ? ) for 0 or 1 match, and the pipe symbol ( | ) representing an OR. Matches can be performed three ways: by specifying a specific string of text, a previously assigned annotation, or a test on an annotation’s attributes. JAPE also includes the ability to define macros that allow for common patterns to be used in multiple rules. The RHS of a JAPE rule is used to create annotations and manipulate the annotation’s attributes. This can be done one of two ways. The simple method involves creating an annotation and then setting the available attributes and each attribute’s value. Although sufficient in many cases, the JAPE syntax cannot handle complex annotation manipulations. For more complex manipulations, Java code may be used on the RHS.

The following is an example of how RAE uses JAPE rules for landmark extraction. One type of indoor landmark that has been identified in a set of route descriptions collected from blind individuals is rooms. Rooms serve as starting locations, ending locations of routes, and as landmarks to be noticed while traveling routes. Therefore, rooms are a desirable
landmark to extract from route directions. One pattern for identifying rooms that regularly appear in the collected texts is the string “room” followed by a number, e.g., “room 405.” In the USU CS department, some rooms’ names also use the pattern of a number, a dash, and then a letter, e.g., “room 401-E.”

To extract these type of landmarks, the string “room” is identified first. A macro is a convenient way to do this:

Macro: ROOM

( {Token.string == "Room"} | {Token.string == "room"} )

This macro, named ROOM, matches a Token with the string attribute equal to “room” or “Room.” This macro can then be used in a rule such as:

Rule: NumberedRoom

( ROOM
  {Token.kind == number}
  ( {Token.string == "-"}
    {Token.orth == upperInitial, Token.length == "1"}
  )?
):numberedroom

--> :numberedroom.Room = {rule="NumberedRoom", kind=room, type=numbered}

This matches the string “room” or “Room,” as defined in the macro ROOM, followed by a number. It also matches an optional dash followed by an upper case letter string of length
one. Any text that matches this rule will cause an annotation named Room to be created with the attributes rule, kind, and type. The attribute names are not predefined by JAPE, but instead are defined and set by the knowledge engineer, allowing for domain dependent annotations and attributes.

Of course, there are other types of rooms in buildings. For example, the pattern of a person’s name followed by the string “office” is common as in “John’s office.” The following JAPE rule can be used to detect this type of natural language pattern:

```
Rule: OfficeRoom
Priority: 10
(
  {Person}
  {Token.category == POS}
  {Token.string == "office"}
):officeroom
-->
:officeroom.Room = {rule="OfficeRoom", kind=room,
                     type=office}
```

This rule matches any person, found by a previously defined JAPE rule responsible for locating general mention of people’s names. The Person annotation must be followed by a token with the category POS, which identifies the possessive string “’s”. Finally, the pattern ends with the string “office.” When matched, a Room annotation is created with attributes rule, kind, and type. Although there are two annotation rules for Room, the annotation’s attributes can be used to determine which rule caused which annotation to be created.

When run against the following natural language text:

```
Room 402 is to your right and room 401-F is to your left. Dr Kulyukin’s office is next to room 402.
```
the previous room-oriented JAPE rules would find and create the following four annotations:

- Room \{kind=room, rule=NumberedRoom, type=numbered\} for the text “Room 402”
- Room \{kind=room, rule=NumberedRoom, type=numbered\} for the text “room 401-
  F”
- Room \{kind=room, rule=OfficeRoom, type=office\} for the text “Dr Kulyukin’s
  office”
- Room \{kind=room, rule=NumberedRoom, type=numbered\} for the text “room 402”

Although it is possible to automate rule development and entity extraction using ma-
chine learning techniques [75], successful training requires a large training corpus. Currently,
there is a small corpus of route directions, 52 indoor route descriptions and 52 outdoor route
descriptions. Thus, RAE’s JAPE rules have been developed manually with the hope that,
as the corpus grows, it may be possible to investigate machine learning techniques for rule
creation.

5.2 Autotagging Details

When a route description is passed to the autotagging process, it is sent through these
processing resources in the following order:

1. ANNIE’s tokenizer
2. ANNIE’s gazetteer
3. RAE-specific gazetteer
4. ANNIE’s sentence splitter
5. ANNIE’s POS tagger
6. ANNIE’s NE transducer with RAE-specific rules
In the tokenizer, all tokens in the route description are annotated. Next, the description is passed to ANNIE’s default gazetteer. The gazetteer provides predefined look-up lists for entities, such as numbers, ordinals, person names, and titles. The text is then passed to the RAE-specific gazetteer, which handles known landmarks from the landmark hierarchy. After the gazetteers, the sentence splitter annotates the boundaries of sentences, and in the POS tagger, each word is assigned a part-of-speech.

The RAE-specific gazetteer is a slightly modified version of ANNIE’s gazetteer. By default, when ANNIE’s version matches strings and words for different look-ups and the possible matches overlap, multiple annotations may be created. For example, when processing the string “John Angus Nicholson” for names, ANNIE’s version would create three annotations: “John Angus Nicholson,” “Angus,” and “Nicholson.” This behavior was modified in the RAE-specific gazetteer to only return the longest single match for these cases, i.e., “John Angus Nicholson.” ANNIE’s original version was retained since it contains a number of gazetteer entries that are useful.

The final step occurs in the NE transducer, which uses RAE-specific JAPE rules. The rules are organized into a series of phases:

1. Setup
2. Entity recognition
3. Noun phrase identification
4. Landmark identification
5. Cleanup

5.2.1 NE Setup Phase

The rule sets associated with the setup phase perform two main functions. The first rule set reclassifies the ANNIE-created POS annotations for words classified as nouns that should have been classified as verbs. The remaining rule sets in this phase identify key words often associated with landmarks.
Some verbs were often marked as nouns by the part-of-speech tagger. It is speculated that this happens because ANNIE’s POS tagger was not trained on a text corpus that included many route descriptions and that people writing the route descriptions did not always use correct grammar or write complete sentences. Patterns were written that identified these instances and reclassified them as verbs. The specific verbs identified in the training set of route descriptions that had instances that were incorrectly classified as nouns were:

approach, arrive, avoid, bear, board, catch, climb, come, continue, count, cross, detect, direct, enter, exit, face, feel, find, follow, get, go, head, hear, leave, look, listen, locate, make, move, navigate, open, pass, press, proceed, push, reach, select, shoreline, smell, step, stay, take, tap, use, trail, turn, use, veer, walk

If one of these words was classified as a noun, the rule set reclassified it as a verb if one of the following cases held true.

1. If the word was the first word in a sentence. For example, “Climb the stairs.”

2. If the word followed a comma and preceded a determinant or the beginning of a prepositional phrase. For example, in “Go up the stairs, take a right.” the word “take” would be classified as a verb.

3. If the word is preceded by the word “to,” as in “You want to climb the stairs.”

4. If the word is preceded by the string “you” and followed by a determinant or the beginning of a prepositional phrase. For example, “You cross the hall.”

5. If the word is preceded by a comma and followed by another noun, for example, in “Turn right, cross Pike St.” the word “cross” would be reclassified.

6. If the word is preceded by “You” and a modal verb such as “will,” for example, “You will cross the street.”

7. If the word is followed by a number, for example, “Then cross 5 streets.”
Another function of the setup process is to identify key words and phrases. One set of words were ordinal words, e.g., “1st,” “second,” and “2nd.” The remaining rule sets in the setup phase are responsible for identifying sets of words that often indicate the presence of landmarks.

Three rule sets - simple transitive, intransitive, and compound transitive - are based off of Jackendoff’s language analysis [53]. Jackendoff’s analysis identifies prepositional words and phrases that are often used in spatial context. A subset of the prepositions has previously been used in landmark classification [32]. In this work, however, the entire set of prepositions are used.

The remaining rule sets were built after observing word usage in the training set of route directions. A summary of the rule sets is shown here, and the complete set of words that each rule set identifies, including the prepositions identified by Jackendoff, are shown in Appendix B.

- Verbs - Annotates the verb forms for “to be” and “to have.” These verbs are annotated because they are frequently used, and annotating them simplifies later rules.
- Cardinal directions - Annotates various forms of cardinal directions, such as “north,” “southwest,” and “SE.”
- Distance - Annotates terms and phrases related to distance, such as “feet,” “yards,” and “meters.”
- Simple Transitive - Annotates simple transitive prepositions. These prepositions, e.g., “above,” “after,” and “at,” are prepositions that represents a spatial relationship with an object.
- Intransitive - Annotates intransitive prepositions, such as “afterward,” “backwards,” and “there.”
- Compound Transitive - Annotates compound transitive phrases, such as “in back of,” “on top of,” and “to the right of.”
• Angle - Annotates short phrases referring using the terms “angle” and “degrees.” Longer phrases including these terms are included as well, e.g., “hard right angle” and “45 degrees.”

• Biased part - Annotates words and phrases which refer to parts of landmarks. For example, “end” is often used in phrases such as “the end of the hall,” or “front” as in “the front of the building.”

• Egocentric reference - Annotates words referring to the traveler himself. Words referring to “guide dog” and “white cane” are annotated as well.

5.2.2 NE Entity Recognition Phase

After the setup phase, the entity recognition phase annotates words and phrases that belong to specific types of landmarks. The idea behind entities is that there are groups of landmarks that are important and have naming patterns that can be identified. Currently, two groups of entities are identified: entrances and streets.

Entrances are any entity that a person can walk through, such as a “door,” “gate,” or “opening.” Often people refer not to the entrance itself, but to a component of the entity, such as a “door knob” or “door jamb.” Entrances often have descriptive adjectives such as “open,” “closed,” and “revolving.” When the entire phrase representing an entrance is found, e.g., “open gate” or “metal door knob,” it is annotated as an EntityEntrance, which represents a landmark.

The second entity, streets, is identified by terms such as “street,” “driveways’, “avenue,” and “alley.” These terms can have various modifiers, such as “intersecting” and “two-way.” They can also have proper names which are identified by ANNIE’s POS tagger, allowing phrases such as “Main Street” to be identified. It is also common to identify mention of streets from the verb “cross” as in “cross Main Street.” When the entire phrase representing a street is found, it is annotated as an EntityStreet, which represents a landmark.
5.2.3 NE Noun Phrase Identification Phase

The noun phrase identification phase purpose is to identify all noun phrases. Since landmarks are objects and are referred to using nouns, every noun phrase is marked. Of course, not all nouns refer to landmarks. The final decision as to whether a noun phrase refers to a landmark is made in the next phase, the landmark identification phase.

The first step of noun phrase identification is to identify phrases that tend to refer to groups of landmarks. Words identifying sets are first marked, e.g., “set,” “bank,” and “couple.” Next phrases using these words are annotated. The phrase may start with adjectives or numbers and will end with the word “of.” Examples include “Three sets of” and “couple of.”

The majority of the work in this phase is spent identifying noun phrases. The output from ANNIE’s POS tagger is used extensively. In addition to nouns, noun phrases often begin with one or more of the following modifiers:

- Proper names, as in “John’s lab.”
- Cardinal directions, as in “north exit.”
- Group phrases, as in “set of stairs” and “bank of elevators.”
- Numbers, as in “two or three steps.”
- Ordinals, as in “second door.”
- Adjectives, as in “long hall.”
- Phrases including “kind of” and “sort of,” as in “kind of a grassy area.” Used when people are unsure of what they are describing.

The rule set has a number of rules that designate specific words and phrases to ignore. This is necessary because some noun phrases are clearly not landmarks. For example, if a person wrote the sentence “Walk the length,” instructing the traveler to walk the length of a previously mentioned landmark such as a hall, the string “length” would be identified as
a noun by the POS tagger. The noun phrase rule has a rule that identifies this situation and prevents “length” from being annotated as a noun phrase. Other words and phrases that are ignored are egocentric references, known landmarks from the landmark hierarchy already annotated in the gazetteer phase, and the word “turn” when it used as a verb.

The remaining rules are the rules that actually annotate noun phrases. In general, noun phrases are preceded by zero or more modifiers and then a series of one or more nouns. There are two special cases. The first contains a set of phrases - “stop sign,” “traffic lights,” and “trash can.” The words in these phrases were not always identified correctly by the POS tagger, making it difficult for them to fit in the general rules. The second exception was the word “number” followed by a sequence of one or more digits, as in “When you get to number 3.”

5.2.4 NE Landmark Identification Phase

The landmark identification phase determines whether the noun phrases annotated in the previous phase are landmarks. Landmarks are rarely mentioned in isolation (although that can happen, and is the reason for the entity recognition phase). Most landmarks are mentioned as part of a phrase. The rule sets in this phase each target one specific type of phrase, and except for the last phrase, phrase secondary landmark, these rule sets are independent of one another.

Phrase Spatial Angle

This rule set contains one rule that finds occurrences of noun phrases followed by a prepositional phrase starting with a simple transitive and ending with mention of an angle. An example would be “The door at an angle to you.” The string “door” in this case would be annotated as a PhraseSpatialAngle and would represent a landmark. This rule was created due to a special case found in a single route description in the training data, but ultimately was not found in any other examples.
Phrase Spatial Simple Transitive

This rule set contains three rules that are centered around simple transitives. The first rule identifies phrases that begin with a simple transitive and end in a noun phrase, e.g., “next to the stairs” would have “stairs” annotated. The second rule is a longer version of the first rule. It identifies a noun phrase immediately before a simple transitive phrase ending in a second noun phrase. There may or may not be a form of “to be” separating the two. For example, in both “The door after the table” or “the door is after the table,” the strings “door” and “table” will be annotated as a landmark. The final rule identifies noun phrases followed by a simple transitive which is then followed by a personal pronoun. Again, there may or may not be a form of “to be” after the noun phrase. For example, in the text “Bob’s office is before hers,” the string “Bob’s office” would be annotated. No matter which rule creates the annotation, the annotation created by this rule set is a PhraseSpatialSimpleTransitive.

Phrase Spatial Compound Transitive

This rule set contains three rules that are centered around compound transitives. The first rule is the simplest. It identifies noun phrases followed by a compound transitive with an optional form of “to be” separating the two. For example, in the text “The office is in between,” the text “office” would be annotated as a landmark. The second rule identifies the reverse situation in which the compound transitive comes first, followed by the noun phrase, with an optional “to be” between them. For example, in the text “in front of the water fountain,” the text “water fountain” would be annotated as a landmark. The third rule extends the first two rules to the case in which two noun phrases are separated by a compound transitive and optional “to be” forms. For example, in the text “the elevators are to the right of the large plant,” the two strings “elevators” and “large plant” would each be annotated as landmarks. In all cases, the annotations created by this rule set are all instances of PhraseSpatialCompoundTransitive.
Phrase Spatial Intransitive

This rule set contains two rules that are centered around intransitives. The first rule identifies noun phrases followed by an intransitive with an optional “to be” form separating the two. For example, in the text “The stairs are afterward,” the text “stairs” would be annotated as a landmark. The second rule identifies the reverse situation in which the intransitive comes first, followed by the noun phrase. In this case, “to be” did not appear to be necessary as people did not use that form in the training set of route descriptions. An example of this rule occurs in the text “and forward of that door.” The text “door” would be annotated as a landmark. Both rules generate PhraseSpatialIntransitive annotations.

Phrase Spatial Distance

This rule set contains three rules centered around mentions of distances. The first rule identifies mention of a distance followed by a noun phrase. The noun phrase may or may not be in a prepositional phrase. For example, in the text “10 feet to the door,” the string “door” would be annotated. The second rule is the reverse situation in which the noun phrase comes first, followed by the distance. These may be separated by an optional form of “to be” and the mention of distance may or may not be in a prepositional phrase. For example, in the text “The street is five feet away,” the string “street” would annotations as a landmark. The final rule is for when two noun phrases are separated by mention of a distance. The presence of “to be” forms is optional and the distance and second noun phrase may or may not be in a prepositional phrase. For example, in the text “The elevator is a few feet from the water fountain,” both “elevator” and “water fountain” would be annotated. The three rules all generate PhraseSpatialDistance annotations.

Phrase Biased Part

This rule set contains three rules centered around mentions of a biased-part phrase. The first rule identifies a noun phrase followed by a biased part. For example, in the text “the door at the end,” the string “door” would be annotated. The second rule is the reverse situation wherein the biased phrase occurs before the noun phrase. The noun phrase may
or may not be in a prepositional phrase. For example, in the text “at the top of the stairs,” the string “stairs” would annotated as a landmark. The final rule is for when two noun phrases are separated by mention of a biased part phrase. The second noun phrase may or may not be in a prepositional phrase. For example, in the text “The door at the top of the stairs,” both “door” and “stairs” would be annotated. The three rules all generate PhraseBiasedNoun annotations.

**Phrase Egocentric Reference**

This rule set contains four rules centered around mentions of an egocentric reference to the traveler or the traveler’s navigation tools. The first rule identifies an egocentric reference, followed by a verb, followed by a noun phrase. For example, in the text “Tell your guide dog to find the building entrance,” the string “building entrance” would be annotated as a landmark. The second rule also starts with an egocentric reference followed by a form of “to be,” followed by a noun phrase which may or may not be in a prepositional phrase. For example, in the text “You are in the front office,” the text “front office” would be annotated. The third rule identifies a noun phrase followed by a comma, followed by an egocentric reference in a prepositional phrase. For example, in the text “The door, to your right,” the string “door” would be annotated. The final rule is similar to the third rule except it does not require a comma, allows for an optional form of “to be” and other verbs. For example, in the text “The lab will be on your left,” the string “lab” will be annotated. All four rules generate the annotation type PhraseEgocentricReference.

**Phrase Verb**

This rule set contains rules that are centered around verbs. Verbs are primarily identified through ANNIE's POS tagger, although there is some correction that takes place during the setup phase as mentioned before. There is some overlap with the phrase egocentric reference rule set in that the string “you” is used in the rules. The difference is that this rule set does not consider the verb “to be.”

The first rule is a verb followed by a noun phrase, often used as a command. For
example, in the text “Open the door,” the string “door” is annotated. The second rule identifies noun phrases followed by a form of “that you,” followed by a verb. For example, in the text “The stairs that you just climbed,” the string “stairs” would be annotated. The third rule is shorter in that it identifies a noun phrase followed by a verb. For example, in the text “The door will close,” the string “door” would be annotated. The next rule identifies two noun phrases separated by a verb. For example, in the text “the men’s room is the second door,” the strings “men’s room” and “second door” would be annotated. The next rule identifies noun phrases that follow variations of “you,” forms of “to have,” and another verb. For example, in the text “You have entered the main hall,” the text “main hall” would be annotated. All rules in this set generate PhraseVerb annotations.

Phrase Secondary Landmark

The phrase secondary landmark is unique among the phrase rules in that it is not independent of the other phrase rules. Instead, it relies on the annotations generated by the other phrase rules. Its purpose is to identify landmarks that are mentioned in conjunction with another landmark but have not been identified in the other phases. The key to identifying the extra landmark are one of the two words “for” or “of.”

The first rule identifies landmarks, followed by “for” or “of,” followed by a noun phrase. For example, in the sentence “You will come right to the doors of the asthma and allergy clinic,” the text “door” would have been annotated earlier as a landmark of PhraseSpatialSimpleTransitive. The rule in this set would then identify the string “asthma and allergy clinic.” The second rule identifies the reverse situation when the noun phrase is followed by a previously annotated landmark. For example, in the text “where the door for the office is located,” the string “office” would be annotated earlier as a PhraseVerb. The rule would then annotate the noun phrase “door” as a landmark. In both cases, the rules generate PhraseSecondaryLandmark annotations.

5.2.5 NE Cleanup Phase

The cleanup phase consists of three rule sets that remove annotations that could cause
confusion. The first rule set removes overlapping annotations. Many of the earlier rule sets are independent of one another. This means that one piece of text may generate multiple annotations. For example, “go through the glass door” may result in two overlapping annotations for “glass” and ”glass door.” In these cases, the annotation inside the longer annotation is deleted - only “glass door” would be reported. Different annotations that match the same text are not deleted. For example, “glass doors” may be annotated as both an EntityEntrance and a PhraseVerb. Since they refer to the exact same text, neither is deleted. Only shorter annotations contained in longer annotations are deleted.

The next cleanup rule set removes references to landmarks which do not exist. These are identified by the negative keywords such as “not,” “no,” and “without.” For example, in the statement “There is no door here,” the string “door” would be annotated as an EntityEntrance. However, the “no” signals that one should not expect a door so in reality it is a non-existent landmark. The EntityEntrance annotation would be deleted.

The final cleanup rule set removes temporary annotations that may have been created by any of the previous rule sets in any of the phases. This helps to reduces the annotation list to only the necessary annotations, most notably the annotations referring to landmarks.

5.3 Experiments and Results

This section reports on the experiments used to evaluate RAE. First, the data set used to evaluate the autotagging process is described. Next, the results of analyzing the data set is described as well as the performance of the autotagging process.

5.3.1 Route Survey

In 2007, an online web-based survey was created and deployed with the purpose of collecting real-world route descriptions from visually-impaired individuals. The survey’s URL was advertised through the e-mail channels of the USU Center for Persons with Disabilities and the National Federation of the Blind (NFB) Utah Chapter in Salt Lake City, Utah. Participation in the survey was completely voluntary. The website did not collect
any identifying information and did not use cookies to track users. Respondents received no compensation, monetary or otherwise, for their participation.

The survey consisted of two sections. The first section collected demographic information consisting of gender, age, education level, level of blindness, the number of years the vision loss had impacted navigation ability, primary navigation aid, whether or not the respondent had received O&M training, navigation skill level, and the presence of other disabilities in addition to visual impairment that could affect navigation. The rating for the navigation skill level was subjective in that it asked respondents to rate their own skill level on the scale from 1 to 5, with 1 being poor and 5 being excellent.

The survey’s second section solicited two route descriptions from respondents. The instructions first asked respondents to describe an outdoor route that could be used to guide a fellow traveler from the entrance of one building to the entrance of another building. Respondents were then asked to describe an indoor route that could be used to guide a fellow traveler from one room in a building to another room in the same building. In both cases, the instructions required that respondents describe real-world routes with which they were familiar. Respondents were also instructed to write the route descriptions as if they were describing the route to a fellow traveler with the same visual impairments and the same traveling experience and skills. For example, respondents who used a guide dog were asked to write their route descriptions so that another guide dog handler would be able to follow the directions. When writing the route descriptions, respondents were asked to assume that the other traveler had no current knowledge of the route they were describing.

There were 52 responses to the survey, providing 104 route descriptions: 52 indoor route descriptions and 52 outdoor route descriptions. The demographics of respondents are summarized in Table 5.1. Eleven respondents reported having another impairment that affected their navigation skills. The breakdown in the additional impairments was:

- Five reported having hearing problems.
- One reported hypopituitarism, a disease of the pituitary gland causing symptoms such as fatigue and muscle weakness.
• One reported problems crossing streets due to post-traumatic stress disorder after an auto-pedestrian accident.

• One reported an inner ear balance disorder that affects mobility and travel skills.

• One reported mobility and gait problems.

• One reported balance problems.

• One reported juvenile rheumatoid arthritis, which caused the respondent to have one fused knee, several joint replacements and limited range of motion in all joints.

The route descriptions were all written in English. When analyzing the routes, no attempt was made to correct any perceived mistakes in the text. Thus, no spelling, punctuation, or grammatical errors were corrected; all route description text was used as originally written by each respondent. The following is an example indoor route description provided by one of the respondents:

Once you are in the lobby of the building, go through security turnstile. Turn left just past turnstile to bank of elevators. Press “up” button (only two to choose from). Take any of the elevators to the third floor. Elevator buttons are labeled with raised numbers. Elevator dings for each floor. When elevator door opens, check the raised number on the outside of the elevator. Turn left to 4th floor on left.

This description also contains examples of the errors encountered in some of the route descriptions. There is an error in the first sentence at the text “in the.” This was probably a simple typing mistake and the respondent most likely meant “in the.” The second error is a missing “the” in the phrase “past turnstile.” The third error is in the fourth sentence, the word “elvators” is misspelled. While the majority of these errors could have been corrected, they were retained since RAE would have to handle text from any user with access to a CRISS website. No assumption was made that users will always go through the process
Table 5.1. Demographics of Route Description Survey Respondents.

<table>
<thead>
<tr>
<th>Field</th>
<th>Response</th>
<th>Number of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>22</td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$age &lt; 20$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$20 \leq age &lt; 30$</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>$30 \leq age &lt; 40$</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>$40 \leq age &lt; 50$</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>$50 \leq age &lt; 60$</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>$60 \leq age &lt; 70$</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$70 \leq age &lt; 80$</td>
<td>2</td>
</tr>
<tr>
<td>Highest Education Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High School</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Some College</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Two-year College</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Undergraduate Degree</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Graduate Degree</td>
<td>23</td>
</tr>
<tr>
<td>Level of Blindness</td>
<td>Complete</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Low-level</td>
<td>24</td>
</tr>
<tr>
<td>Number of Years</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$years &lt; 10$</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>$10 \leq years &lt; 20$</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>$20 \leq years &lt; 30$</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>$30 \leq years &lt; 40$</td>
<td>11</td>
</tr>
<tr>
<td>Vision Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$40 \leq years &lt; 50$</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$50 \leq years &lt; 60$</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>$60 \leq years &lt; 70$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$70 \leq years &lt; 80$</td>
<td>1</td>
</tr>
<tr>
<td>Navigation Skill Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Very Good</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
<td>15</td>
</tr>
<tr>
<td>Navigation Aid</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cane</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Guide Dog</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>Received O&amp;M training</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Has Other Impairment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>11</td>
</tr>
<tr>
<td>Affecting Navigation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>41</td>
</tr>
</tbody>
</table>
of correcting the route description text with tools such as spelling and grammar checkers before submitting their route description.

5.3.2 Evaluating RAE’s Landmark Autotagging

Each route description was randomly placed into either a training set or an evaluation set. Two-thirds of the route descriptions were placed in the training set, and the remainder were placed in the evaluation set. Table 5.2 shows the number of route descriptions in each set and how the descriptions were spread across the sets. Although each respondent submitted both indoor and outdoor route descriptions, the two descriptions were placed into sets independent of one another. It was possible, for example, that the indoor route description was placed in the training set and the outdoor route description for the same respondent was placed in the evaluation set.

The training set was used to develop text patterns that could be used as a basis for the JAPE rules. The landmarks were manually identified in each sentence, and the set of the autotags that RAE was expected to generate was manually indicated. The JAPE rules were built based on this expected output. When the autotagging rules were complete, they were tested using the evaluation set of route directions. In the evaluation set, all landmarks in each route description were manually identified. The autotagging process was then run, and the set of extracted landmarks were compared against the manually identified set of landmarks.

The process of analyzing the text also yielded information about the route descriptions. Table 5.3 shows that in both the training set and the evaluation set, the outdoor route de-
Table 5.3. Sentence Counts Per Route Description.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Total Sentence Count</th>
<th>Average # of Sentences per Description</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>352</td>
<td>10.4</td>
<td>6.68</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>686</td>
<td>20.2</td>
<td>18.15</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>164</td>
<td>9.1</td>
<td>4.03</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>276</td>
<td>15.3</td>
<td>8.41</td>
</tr>
</tbody>
</table>

Table 5.4. Word Counts Per Route Description.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Total Word Count</th>
<th>Average # of Words per Description</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>5,459</td>
<td>160.6</td>
<td>118.58</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>10,329</td>
<td>303.8</td>
<td>307.76</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>2,271</td>
<td>126.2</td>
<td>68.32</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>4,090</td>
<td>227.2</td>
<td>141.17</td>
</tr>
</tbody>
</table>

Descriptions contained more sentences than the indoor route descriptions. For the purpose of this discussion, a sentence is defined to be a string of text annotated by ANNIE’s sentence splitter as a `Sentence` annotation. It should be noted that this definition does not always lead to what could be considered a valid sentence in terms of English grammar and punctuation rules. This is due to the fact that, as mentioned earlier, several respondents did not always use correct punctuation and grammar, which sometimes caused the sentence splitter to mark or miss sentence boundaries that a human may notate differently.

Since outdoor route descriptions have more sentences, it follows that they also contain more words. Table 5.4 shows word counts on a per route description basis. Again, for both the training and the evaluation sets, the outdoor route descriptions contain more words than the indoor route descriptions in the set. However, when looking at word counts on a per sentence basis (see Table 5.5), the average number of words per sentence remains consistent across all descriptions. The higher count of words in the outdoor route descriptions is more
Table 5.5. Word Counts Per Route Sentence.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Total Word Count</th>
<th>Average # of Words per Sentence</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>5,459</td>
<td>15.51</td>
<td>11.50</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>10,329</td>
<td>15.06</td>
<td>8.70</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>2,271</td>
<td>13.85</td>
<td>7.69</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>4,090</td>
<td>14.82</td>
<td>10.15</td>
</tr>
</tbody>
</table>

Table 5.6. Landmark Counts Per Route Description.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Total Landmark Count</th>
<th>Average # of Landmarks per Description</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>714</td>
<td>21.0</td>
<td>14.99</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>1,362</td>
<td>40.1</td>
<td>34.36</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>301</td>
<td>16.7</td>
<td>8.63</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>565</td>
<td>31.4</td>
<td>19.10</td>
</tr>
</tbody>
</table>

A function of the number of sentences and not the sentence length.

The outdoor route descriptions’ additional length may be a result of the additional complexities of following routes in outdoor environments. Outdoor routes can be longer than indoor routes, and travelers may require more information along the routes in order to maintain orientation. Another possible reason is that outdoor environments are less structured than indoor environments. Less structure may lead to more description in order to help resolve potential ambiguity when a traveler is sensing the environment. Outdoor routes also have elements of safety, e.g., crossing streets, that do not exist when following indoor routes. In order to ensure travelers’ safety, more details need to be added describing the potential problems.

When the number of landmarks is analyzed in the route descriptions, the patterns are similar to the sentence and word counts. As with the sentences and words per route description, the number of landmarks mentioned in outdoor route descriptions (see Table 5.6)
Table 5.7. Landmark Counts Per Route Sentence.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Total Landmark Count</th>
<th>Average # of Landmarks per Sentence</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>714</td>
<td>2.0</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>1,362</td>
<td>2.0</td>
<td>1.44</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>301</td>
<td>1.8</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>565</td>
<td>2.0</td>
<td>1.62</td>
</tr>
</tbody>
</table>

Table 5.8. Counts of Sentences without Landmark Per Route Description.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Total # of Sentences with no Landmark</th>
<th>Average # of Sentences with no Landmark per Description</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>27</td>
<td>.8</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>55</td>
<td>1.6</td>
<td>2.58</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>9</td>
<td>.5</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>25</td>
<td>1.4</td>
<td>1.50</td>
</tr>
</tbody>
</table>

is approximately double the number mentioned in indoor route descriptions. Likewise, the number of landmarks per sentence appears to be consistent with approximately two different landmarks mentioned on average (see Table 5.7). The number of landmarks remains consistent on a per sentence basis, following the pattern seen in the words per sentence. Assuming that outdoor route descriptions are longer due to their length, complexity, and increased set of safety issues, it would follow that more landmarks would need to be mentioned in order to allow someone to successfully navigate the described route.

Not all sentences contain references to landmarks. Tables 5.8 and 5.9 show that while most sentences mention at least one landmark, between 5% and 10% of the sentences do not. These sentences often consist of short simple commands, e.g., “Turn right” and “Cross straight ahead.” Other sentences without landmarks add commentary and more description, e.g., “A white cane is especially useful here” and “Be careful here, it’s not the easiest place
Table 5.9. Ratios of Sentences without Landmark Per Route Description.

<table>
<thead>
<tr>
<th>Set Name</th>
<th>Descriptions’ Location</th>
<th>Average % of Sentences with no Landmark per Description</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>Inside</td>
<td>7.3%</td>
<td>9.90</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>6.9%</td>
<td>6.87</td>
</tr>
<tr>
<td>Evaluation Set</td>
<td>Inside</td>
<td>4.9%</td>
<td>6.31</td>
</tr>
<tr>
<td></td>
<td>Outside</td>
<td>8.8%</td>
<td>9.60</td>
</tr>
</tbody>
</table>

to navigate.” There are also sentences specifically mentioning the lack of landmarks, e.g., “There are no steps or uneven spots” and “There is no light so listen well.” These sentences extend the information about the landmarks mentioned in the sentences before and after. Landmarks are involved, but the sentences, when referenced as standalone units or thoughts, do not directly reference any landmark.

After building the JAPE rules based on the training set of route descriptions, the rules were evaluated using the evaluation set in which the landmarks had been manually identified. The evaluation set was run against the JAPE rules, and the following four scores were calculated:

1. Correct - signified that the autotagged and manually annotated landmarks match exactly.

2. Partial - signified that the autotagged landmark overlaps the manually annotated landmark and is not an exact match. For example, in the text “From the information desk walk straight,” the string “information desk” was manually annotated, but the autotagging process extracted the longer string “information desk walk.”

3. Missing - signified that a landmark was manually annotated, but was not found during the autotagging process.
Table 5.10. Results of Autotagging on Evaluation Route Descriptions.

<table>
<thead>
<tr>
<th></th>
<th>Inside Route Descriptions</th>
<th>Outside Route Descriptions</th>
<th>All Route Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manually Annotated</td>
<td>301</td>
<td>565</td>
<td>866</td>
</tr>
<tr>
<td>Correct</td>
<td>252</td>
<td>460</td>
<td>712</td>
</tr>
<tr>
<td>Partial</td>
<td>18</td>
<td>28</td>
<td>46</td>
</tr>
<tr>
<td>Missing</td>
<td>31</td>
<td>77</td>
<td>108</td>
</tr>
<tr>
<td>Incorrect</td>
<td>27</td>
<td>102</td>
<td>129</td>
</tr>
</tbody>
</table>

4. Incorrect - signified that the autotagging process incorrectly identified some text segment as a landmark.

These scores were then used to calculate precision \((P)\), recall \((R)\), and F-measure \((F)\) following standard information extraction evaluation metrics [11,84] for the indoor route descriptions, the outdoor route descriptions, and the all the route descriptions from the evaluation set. The scores were calculated using the standard formulas:

\[
P = \frac{\text{correct} + (\text{partial} \times .5)}{\text{correct} + \text{partial} + \text{incorrect}}
\] (5.1)

\[
R = \frac{\text{correct} + (\text{partial} \times .5)}{\text{correct} + \text{partial} + \text{missing}}
\] (5.2)

\[
F = \frac{(\beta^2 + 1.0) \times P \times R}{\beta^2 \times P + R}
\] (5.3)

The \(\beta\) value in the F-measure is the relative importance given to recall over precision. Here, the two measures were deemed equally important so \(\beta\) was set at 1.0. The results for the correct, partial, missing, and incorrect values are shown in Table 5.10, and the results for the precision, recall, and f-measure scores are shown in Table 5.11.

The results show that the majority of the extracted landmarks are relevant, with \(P = 0.8286\) for all route descriptions. Likewise, it was found that the majority of the
Table 5.11. Computed Scores for Each Evaluation Set.

<table>
<thead>
<tr>
<th></th>
<th>Inside Route Descriptions</th>
<th>Outside Route Descriptions</th>
<th>All Route Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.8788</td>
<td>0.8034</td>
<td>0.8286</td>
</tr>
<tr>
<td>Recall</td>
<td>0.8671</td>
<td>0.8389</td>
<td>0.8486</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.8729</td>
<td>0.8208</td>
<td>0.8386</td>
</tr>
</tbody>
</table>

expected landmarks were annotated, with $R = 0.8487$ for all route descriptions. When the evaluation set is broken into inside and outside route directions, it can be seen that all three measures, $P$, $R$, and $F$, are higher for inside route descriptions than for the outside descriptions. This may be because indoor environments tend to be more structured than outdoor environments, leading to less variation in the language used to describe landmarks and the actions associated with them.

A number of landmarks were not extracted, 108 or 12.4% for all route descriptions. 129 incorrect landmarks were autotagged. These results are not surprising given the data set. The route descriptions are all in English but have a wide variety styles. Some sentences are simple and declarative, e.g., “Go out the door and make a right.” Others are longer with more complicated phrasing, e.g., “If you miss it there are benches on the outside which will tell you your near where you want to be listen for traffic though because there is a drive through you’ve passed the door if you get to that side of the building and need to backtrack.” Since people have a wide variety of writing styles, it is unlikely that all possible grammatical patterns would have been seen in the 68 training examples used to build the JAPE rules. In order to solve or at least reduce these problems, additional route descriptions need to be collected and analyzed. This will help to improve and refine the current rules, as well as help to identify additional language patterns that could be implemented as new rules.

In addition to the performance, some details emerged about the extracted annotations. Table 5.12 and Figure 5.1 show the counts of the different types of landmark annotations. The totals for the columns in Table 5.12 are not the same as the landmark totals previously
mentioned. This is due to two reasons. First, landmarks can be annotated more than once. For example, in the sentence “Enter the glass door at the end of the hall on your right,” the landmark “glass door” would be annotated two ways by the system. It would be marked as an **EntityEntrance** landmark because it is an entrance and “door” is considered an entrance. It would also be annotated as a **PhraseVerb** landmark because it appears in close proximity to the verb “Enter.” The third annotation it would have would be **PhraseBiasedNoun** since the phrase “at the end” has also been identified as a key marker for landmarks.

The two most used landmark annotations are **PhraseSpatialSimpleTransitive** and **PhraseVerb**. It is not surprising that this is the case. Route directions are a series of instructions in which actions are committed at specific landmarks. Actions are communicated as verbs in the text. People typically do not usually write just “Enter” but will also explain what to enter, “Enter the main door,” designating a landmark with which to commit or associate the action. Simple transitive words, such as “At,” “before,” “to,” and “up,” communicate spatial relationships. Visually-impaired travelers require more explicit spatial relationships in their route directions than a sighted person often does, because they do not have that gestalt view of the world. Whereas a sighted person can make do with an instruction such as “When you see the main doors, enter the building” a visually impaired person may need more information, for example, ”At the bench, turn right. Go up the stairs. Walk forward to the main doors. Enter the building.” The least used landmark annotation was the **PhraseSpatialAngle** with only one instance in both the training set and evaluation set. **PhraseSpatialAngle** may be too specific, at least in this set of route descriptions. More route descriptions are needed to decide whether to expand the scope of this particular annotation or to delete it from the rule set.

The two entity specific annotations, **EntityEntrance** and **EntityStreet**, were each designed to specifically target a certain type of landmark. Table 5.12 and Figure 5.1 both show that each entity annotation is extracted from the routes descriptions. **EntityStreet** is primarily outdoor based, with few hits on indoor routes. Mention of streets does occa-
sionally happen when a route description describes initially entering an indoor environment or leaving an indoor environment, and their location or sound is used as a reference landmark. In the evaluation indoor set, neither entity annotation was responsible for uniquely matching a landmark; they were always part of multiple annotations matching a landmark.

The intention of the entity annotations was to pick stray landmarks that the phrase-based annotations missed. As Table 5.13 shows, this happened in both the training and evaluation sets, but less so in the evaluation set. The table compares single matches and multiple matches. Single matches occur when only the entity annotation matched a landmark. A multiple match is when multiple annotations, one of which was the entity annotation, matched a single landmark. The numbers on the training set suggest that the entity annotations can provide some help. Approximately 11% of the entity annotations found were matched only by the entity annotations in the training set. That count fell to around 3% in the evaluation set. Because the rules were built based on the specific language used in the training set, it performed better. The process works to a degree with the evaluation set, but different terms and language use reduced the effectiveness.
Figure 5.1. Bar chart of counts for different types of landmark annotations.

Table 5.13. Entity Annotation Counts.

<table>
<thead>
<tr>
<th></th>
<th>EntityEntrance</th>
<th>EntityStreet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training Set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indoors</strong></td>
<td>Single</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>127</td>
</tr>
<tr>
<td><strong>Training Set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outdoors</strong></td>
<td>Single</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>92</td>
</tr>
<tr>
<td><strong>Evaluation Set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indoors</strong></td>
<td>Single</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>59</td>
</tr>
<tr>
<td><strong>Evaluation Set</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outdoors</strong></td>
<td>Single</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Multiple</td>
<td>36</td>
</tr>
</tbody>
</table>
5.4 Summary

This work shows that information extraction techniques can be used to process natural language route descriptions and can be used to identify the routes’ landmarks. This process takes advantage of the language and word patterns people use when describing routes. Since route descriptions are generally written in a common manner, even by different writers, these patterns and word usages that exist in the descriptions can be exploited. The most important identifiers for landmarks are verbs and prepositional phrases referring to spatial relationships. Other patterns also exist that can be used to identify landmarks that verb and prepositional phrase patterns miss.

Due to the variety of writing styles and grammatical errors in the text, general patterns are not able to identify all landmarks. Results can be improved by implementing rules targeting specific types of landmarks that have specific naming patterns. The examples used in this research were entrances and streets. In many cases, multiple rules, general and specific, will identify the same landmark. However, more specific rules and entity specific rules can locate landmarks missed by the more general rules. In order to reduce multiple mentions of a landmark in the result set created my multiple rules, a landmark’s location in the text can be easily checked with the other landmarks’ locations, and multiple mentions of a landmark can be reduced to a single annotation.

Improving the quality of the rules is most likely a matter of training on more data. All the rules in this work were built by hand. This was done due to the relatively small number of route descriptions in the corpus. A larger set of rules covering more cases could be built using a larger set of training data. With a sufficiently large set of route descriptions, it may be possible to use machine learning techniques in order to automate the process of building rules.
CHAPTER 6
PATH INFERENCE

6.1 Introduction

According to Golledge’s survey of how the disabled deal with the geography [38], the visually impaired understand the world in terms of routes. A consequence of this type of spatial understanding is that a person with a visual impairment, especially one unfamiliar with a given area, may travel a set of routes without realizing that the routes share common areas and landmarks. This can lead to the person to travel routes that are longer than necessary and to miss potential short cuts. Understanding the spatial relationships between routes can potentially lead to shorter and more efficient paths.

As an example of the problem, consider the following. Suppose a new student at Utah State University (USU) is given the following route description by another student who was already familiar with the campus. The route description describes a route (see Figure 6.1), call it route $R\cdot A$, that guides a traveler from one of the entrances of the Animal Science building, $L\cdot A$ in Figure 6.1, to one of the entrance’s of the Ray B. West building, labeled as $L\cdot C$ in the figure, on USU’s campus. The route passes through an area of USU’s campus known as the Quad, which is a large grassy area with two sidewalks, one running north/south and the other running east/west. The two sidewalks intersect in the middle of the Quad, labeled as $L\cdot B$.

**Description for R-A:** Exit the Animal Science building doors on the south side. Walk straight until you find the sidewalk entrance to the Quad’s sidewalk. Pass the main sidewalk intersection. Walk south until you detect a road and then carefully cross the street. Continue to walk south until you find the doors to the Ray B. West building.
At a later date, the student is given a route description for another route, call it route \( R-B \), which is shown in Figure 6.2. The description describes a route that starts at one of the entrance’s to the Old Main building, labeled \( L-D \) in the figure, to the entrance to the Distance Learning Center offices, \( L-E \), in the Eccles Conference Center. The route passes through the Quad as did route \( R-A \) but by way of the Quad’s other sidewalk.

**Description for R-B:** Exit Old Main walking east. You will walk through the Quad, passing the intersection. Keep walking straight until you run into grass and then turn left, walking north. Walk until you detect the bike racks on your right and then turn right. Walk east until you find the stairs leading to the entrance to the distance learning center.

As Figures 6.1 and 6.2 show, the descriptions for \( R-A \) and \( R-B \) both describe routes that pass through the center of the Quad where the two sidewalks intersect, location \( L-B \). This is useful if a visually-impaired traveler needs to go from Old Main to Ray B. West, but does not know anyone who knows the area and can describe the route. By taking advantage of previous spatial knowledge gained by learning the two route descriptions, a
Figure 6.2. Example route R-B.

traveler can combine the first part of route R-B with the second part of route R-A to form the new route R-C (see Figure 6.3). For a sighted traveler this is trivial, but depending on the level of experience with the USU campus and the Quad, a traveler with a visual impairment, especially someone with total vision loss, may not be aware that the “main sidewalk intersection” mentioned in route R-A and the “intersection” mentioned in route R-B actually refer to the same physical location. If this relationship is not established, the visually-impaired traveler may not be aware that route R-C exists and consists of route segments with which he is already familiar.

It has been shown that RAE’s autotagging process can transform a natural language route description into a structured list of route statements. The autotagging process identifies landmarks in each route statement and tags that route statement with those landmarks. In a CRISS-based system, users would further refine the route by editing the autotagged landmark tags and annotating route statements with additional landmarks. Because the route descriptions are no longer pure free-form text, but instead are contained within a data structure, sets of tagged route description can now be processed using various algorithms.

By taking advantage of the structure imposed on the natural language route description,
RAE is capable of finding new routes that do not yet have route descriptions in the system. It does this, not through natural language processing methods, but by transforming the set of the tagged route descriptions into a graph structure that can then be processed with well-known graph search algorithms.

Path Inference is the process of inferring a new, previously unknown route from a set of previously known routes. In the example, route $R-C$ was inferred from the known routes $R-A$ and $R-B$. This was possible due to the common landmark shared by the known routes. The goal of RAE’s path inference process is to accomplish this task starting from a known set of tagged route descriptions. The end result is a new route description which is not originally part of the original, tagged route descriptions.

RAE’s path inference serves two purposes. First, it allows users to enter sets of route descriptions at a faster rate. Writing route descriptions takes time, and an area such as a university campus requires many route descriptions to be entered in order for the system to be useful. RAE’s path inference speeds up this process since new route descriptions can be generated automatically. The second purpose is that it allows new routes of which users may be unaware to be discovered, increasing users’ spatial knowledge of the area.
To find the new route description, the path process transforms the set of tagged route descriptions into a directed graph, or digraph, with weight edges. Then, when given a starting landmark and an end landmark, the system searches the digraph for a path between the two landmarks. If a path is found, a route description is built and returned to the user. No path being found signifies that a route cannot be discovered in the digraph, but it does not signify that a route does not exist in the real-world. When a path is not found, this is a sign that the system most likely needs more route descriptions.

6.2 Transformation into a Digraph

In more formal terms, the path inference process begins with the transformation of the tagged route statements into a directed graph. The digraph consists of two types of nodes, statement nodes and landmark nodes, and all edges are directed edges. During the transformation, every route description in CRISS’s collection is added to the digraph through a process that maps every route statement and its associated landmark tags to nodes and edges of the digraph.

The graph contains directed edges because there is no guarantee that the same landmarks and statements would be used when describing a route from A to B as when describing a route from B to A. This is due to how some visually-impaired travelers perform trailing, i.e., following the edge of a feature such as a wall or the sidewalk. When walking down a hall in one direction, for example, they may trail the right hand wall and in that instance encounter one set of landmarks. When going in the opposite direction, since they trail the wall on the other side, a different set of landmarks may be encountered.

Each route statement is represented by a statement node. A route statement that does not have a landmark tag associated with it is connected by an edge to the next statement node in the graph. This connection represents a precedence relationship between the two route statements signifying that the action in a route statement cannot be performed before the actions in the preceding route statements have been performed. If a route statement has been associated with one or more landmarks, each landmark is represented in the digraph by a landmark node. The statement node and its landmark nodes are connected by edges
representing an *association relationship* signifying that the route statement was tagged with the landmarks. A precedence edge from the landmark node then connects the landmark node to the next statement node in the description. This connection continues to signify the ordering of the route statements in a route description even when route statements have been tagged. Statement nodes can precede either statement nodes or landmark nodes. Landmark nodes can only precede statement nodes, since some action is required to move from one landmark to another landmark.

As an example, consider the following natural language route description contributed by a visually-impaired USU student. It describes an indoor route that leads a traveler from the Quick Stop, a small room acting as a convenience store, to the Hub, one of the main areas on campus for buying meals. The route occurs entirely within one building, the Taggart Student Center.

> You are standing with your back to the south entrance to the Quick Stop. Turn left so you are walking east. On your left you will pass the ATM machines which make distinctive sounds, and the campus post office and mailbox. You will pass the entrance to the financial aid office on your right and several bulletin boards. Continue walking east and passing offices, the barber shop, and the copy center as you walk down this long hall. Towards the eastern end of the building, you will come to a wide open area on your left. Turn left and walk a little north. Pass Taco Time on your left, and look for a small opening on your lift. This opening will have a cashier counter on your right. Turn left and enter the world of the Hub. You will find a wide variety of food stations around a semicircle.

After processing by the autotagging process and further tagging by users, the beginning of the route description would be represented by the structure shown in Figure 6.4. It is assumed the landmarks used as tags are present in the landmark hierarchy. Once tagged, a route description can be transformed into as part of the digraph. The resulting transformation of the example description is shown in Figure 6.5. It should be noted that the figure contains a pentagon node representing the start of the route, and a connection from
Route ID: R-3
Start Landmark: Landmark(name="QUICK STOP", id="L-500")
End Landmark: Landmark(name="THE HUB", id="L-789")
Statement List

1. Statement
   id: “S-3-1”
   text: “You are standing with your back to the south entrance to the Quick Stop.”
   landmarks: Landmark(name="SOUTH ENTRANCE", id="L-501")

2. Statement
   id: “S-3-2”
   text: “Turn left so you are walking east.”
   landmarks: NIL

3. Statement
   id: “S-3-3”
   text: “ATM machines which make distinctive sounds, and the campus post office and mailbox.”
   landmarks:
      Landmark(name="ATM", id="L-550"),
      Landmark(name="POST OFFICE", id="L-551"),
      Landmark(name="MAILBOXES", id="L-552")

4. Statement
   id: “S-3-4”
   ...

Figure 6.4. Partial example of a tagged route description.

that node to the start landmark representing the route. This node and connection are not actually added to the system’s digraph. Rather the start of the route is represented as an edge weight. The node is added for visual and discussion purposes. The final representation also assigns weights to each edge, but these weights are not displayed in this representation. Edge weights are discussed in detail in Section 6.4.

The algorithm for performing the transformation from a set of landmark tagged route
descriptions to a digraph is shown in Figure 6.6. The graph’s set of vertices, $V$, contains nodes that can reference both route statements and landmarks. All route statements are unique, even if they contain the same text. Therefore, all route statements in every route are represented as a unique node in the graph. Landmarks are represented by a single node, even when used as a tag in different routes. Therefore, there exists at most one node in $V$ for each unique landmark id in the landmark hierarchy.

The loop at line 17 is responsible for connecting a route statement to all its landmark tags. If the route statement does not have any landmark tags, it will simply be connected to the next route statement or end landmark, if it is the last route statement. The check at line 19 prevents an edge from the first route statement to the start landmark from being created. The start landmark is always set as the first node for a route in line 5 and is always connected to the route’s first statement in the for-loop at line 10. The start landmark may or may not be one of the landmarks in the first statement’s set of landmark tags. If it is part of the landmark tag set, an edge will not be created that refers back to the start landmark.

The final check at line 30 determines if the end landmark has been added previously.
Function BUILD_GRAPH(TaggedRouteSet)

1. foreach route ∈ TaggedRouteSet do
2.     startLandmark ← route’s start landmark;
3.     endLandmark ← route’s end landmark;
4.     statements ← route’s list of route statements;
6.     previousNodes ← \{startLandmark\};
7.     for statementId = 1 to length(statements) do
8.         statement ← statements[statementId];
9.         if statementId = 1 then
10.            CONNECT(E[G], previousNodes, statement, true);
11.        else
12.            CONNECT(E[G], previousNodes, statement, false);
13.        end
15.     previousNodes ← \{statement\};
16.     nextNodes ← \emptyset;
17.     foreach landmark ∈ statement’s set of landmarks do
19.         if statementId = 1 and landmark = startLandmark then
20.             nextNodes ← nextNodes \cup \{statement\};
21.         else
22.             nextNodes ← nextNodes \cup \{landmark\};
23.             CONNECT(E[G], previousNodes, landmark, false);
24.         end
25.     end
26.     if nextNodes ≠ \emptyset then
27.         previousNodes ← nextNodes;
28.     end
29. end
30. if endLandmark ∉ statement’s set of landmark tags then
32.     previousNodes ← \{statement\};
33.     CONNECT(E[G], previousNodes, endLandmark, false);
34. end
35. end
36. return G;

Figure 6.6. BUILD_GRAPH() function for transforming a set of tagged route descriptions into digraph $G = (V, E)$. 

Function CONNECT(E, previousNodes, toNode, isFirstStatement)

1. foreach fromNode ∈ previousNodes do
2.   if fromNode is a Statement and toNode is a Statement then
3.     E ← $E \cup$ PrecedenceEdge(fromNode, toNode, TO_STATEMENT_COST); 
4.   else if fromNode is a Landmark and toNode is a Statement then
5.     if isFirstStatement then
6.       E ← $E \cup$ PrecedenceEdge(fromNode, toNode, START_ROUTE_COST); 
7.     else
8.       E ← $E \cup$ PrecedenceEdge(fromNode, toNode, TO_STATEMENT_COST); 
9.     end
10.   else if fromNode is a Statement and toNode is a Landmark then
11.     E ← $E \cup$ AssociationEdge(fromNode, toNode, TO_LANDMARK_COST); 
12.   end
13. end

Figure 6.7. CONNECT() function for creating weighted edges in the digraph.

If it was added, this signifies that the end landmark was one of the landmark’s in the set of tags for the route’s final route statement. If the end landmark was not one of the final route statement’s tags, check will add a connection between the statement and the end landmark.

BUILD_GRAPH() uses the function CONNECT(), shown in Figure 6.7, to create the edges in the digraph. As mentioned earlier, there is no node representing the actual start of the route. All edges have an associated cost or weight. There are four possible weights discussed in Section 6.4. However, it should be noted here that whenever there is a PrecedenceEdge that connects the start landmark of a route to the first route statement, this edge has the cost START_ROUTE_COST. This cost signifies the start of the route and will be higher than the other costs on other edges. For comparison purposes, Figure 6.8 shows the partial route from before as created with edge weights by BUILD_GRAPH() and CONNECT().
6.3 Inferring Paths

Once the routes are transformed into a graph, the graph can be used to find a path from one landmark node to another landmark node. If a path is found, the statement nodes along the path can be combined to create a new route. The new route description returned from the process may assist travelers in recognizing geographic relationships of which they were not previously aware.

The basic process can be illustrated by extending the example from Section 6.1. Figure 6.9 shows the descriptions A and B transformed into a digraph using the ids in Tables 6.1 and 6.2. Route description R-1 describes the route from Old Main (landmark L-1) to the Distance Learning Center (landmark L-2), and route description R-2 describes the route from Animal Science (landmark L-3) to Ray B. West (landmark L-4). If no route description yet exists in the route description database from Old Main to Ray B. West, the path inference process could be used to find the path in the digraph from L-1 to L-4.

To build a new route description, the statements along the path are joined to create a new list of route statements. The original route statements are cloned, including their
Table 6.1. Example Landmark Ids for Descriptions A and B.

<table>
<thead>
<tr>
<th>ID</th>
<th>Landmark Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-1</td>
<td>Old Main</td>
</tr>
<tr>
<td>L-2</td>
<td>Distance Learning Center</td>
</tr>
<tr>
<td>L-3</td>
<td>Animal Science</td>
</tr>
<tr>
<td>L-4</td>
<td>Ray B. West</td>
</tr>
<tr>
<td>L-99</td>
<td>Quad sidewalk intersection</td>
</tr>
</tbody>
</table>

Table 6.2. Example Route Statement Ids for Descriptions A and B.

<table>
<thead>
<tr>
<th>Route</th>
<th>ID</th>
<th>Statement</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>R-2</td>
<td>S-2-1 Exit the Animal Science building doors on the south side.</td>
<td>L-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-2-2 Walk straight until you find the sidewalk entrance to the Quad’s sidewalk.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-2-3 Pass the main sidewalk intersection.</td>
<td>L-99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-2-4 Walk south until you detect a road and then carefully cross the street.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-2-5 Continue to walk south until you find the doors to the Ray B. West building.</td>
<td>L-4</td>
</tr>
<tr>
<td>B</td>
<td>R-1</td>
<td>S-1-1 Exit Old Main walking east.</td>
<td>L-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-1-2 You will walk through the Quad, passing the intersection.</td>
<td>L-99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-1-3 Keep walking straight until you run into grass and then turn left, walking north.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-1-4 Walk until you detect the bike racks on your right and then turn right.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-1-5 Walk east until you find the stairs leading to the entrance to the distance learning center.</td>
<td>L-2</td>
</tr>
</tbody>
</table>
tags, and the new description and its route statements are given a new route id. Route statements are cloned so that the new route description can be edited without affecting the original route descriptions. In the example, the new route description may be given the route id R-3 and its list of route statements would be:

- S-3-1 (cloned from S-1-1): Exit Old Main walking east.
- S-3-2 (cloned from S-1-2): You will walk through the Quad, passing the intersection.
- S-3-3 (cloned from S-2-4): Walk south until you detect a road and then carefully cross the street.
- S-3-4 (cloned from S-2-5): Continue to walk south until you find the doors to the Ray B. West building.

Following this new description would be relatively straightforward since the actions are from known route descriptions previously approved by users. However, there is a problem
with the description. Reexamining Figure 6.3 shows that the new route description requires the traveler to make a right turn at the Quad’s sidewalk intersection. Statement S-3-2 does not mention a turn, but instead appears to instruct a traveler to continue walking straight, “passing the intersection.” This is an example of action inconsistency, the situation in which actions described in two different route descriptions result in inconsistency when the statements are joined in a new route description.

Action consistency is not guaranteed by the path inference process. The process does not check for inconsistencies because the information to make such a decision is not stored in the system. To do so would require that the system perform additional natural language processing of the descriptions, and even then the new processed description still could not guarantee freedom of inconsistencies. To address the problem, RAE again employs users’ spatial knowledge and abilities, similar to ShopTalk, as components of the system.

Once a new route description is found, it is passed back to a user. The user is given the opportunity to read and, if necessary, edit the route for clarity. The review step is also necessary to ensure that the route can be followed safely. Therefore, in a CRISS-based system, the route description must be admitted into the system, i.e., checked by a user familiar with the area to ensure that the description is both understandable and can be used to safely guide a person. Once admitted, the new route description would be made available to all users. In the example, a knowledgeable user would edit the route description, perhaps replacing the phrase “passing the intersection” with an additional route statement, resulting in a new route description that can easily be followed.

- S-3-1 : Exit Old Main walking east.
- S-3-2 : You will walk through the Quad.
- S-3-3 : Turn right at the sidewalk intersection.
- S-3-4 : Walk south until you detect a road and then carefully cross the street.
- S-3-5 : Continue to walk south until you find the doors to the Ray B. West building.
The main purpose of path inference is not to create route descriptions that are completely accurate. No matter what algorithm is employed, route descriptions would have to be checked by knowledgeable travelers for safety. The path inference process is more about helping users learn the spatial relationships among routes with which they are already familiar.

### 6.4 Path Inference Heuristics

As the set of route descriptions grow, the chance that multiple paths existing between pairs of landmarks will grow is very likely. Choosing a “good” path from which to build a new route description requires that choices be made regarding which path should be preferred over others. The path inference process is designed around three simple heuristics that aid the system in choosing a path. Two of the heuristics address action inconsistency, and the third addresses the final length of the path.

Although it is not possible to completely eliminate action inconsistency, it is possible to reduce the likelihood of inconsistencies in the generated route descriptions. This is possible through the application of two simple heuristics. The first heuristic, *heuristic-1*, states that when determining a path in the digraph, choose a path that uses as few route descriptions as possible. The second heuristic, *heuristic-2*, states that when determining a path in the digraph and joining two route descriptions, it is preferable to join route descriptions at their start and ending landmarks rather than at landmarks mentioned in the middle of routes.

Action inconsistencies occur when a route statement from one route description is followed by a route statement from another route description. The purpose of heuristic-1 is to minimize the number of changes from one route to another route when searching for a path in the digraph. Figure 6.10 provides an example of how this heuristic works. Suppose a new route description is needed that instructs a person how to go from landmark L-182 to L-428. The figure’s digraph has two possible paths that could be used. The upper path uses three previously known routes, R-1, R-3, and R-4, to create a new route description, R-5, consisting of the three route statements [S-1-1, S-3-1, S-4-1]. The other potential path uses the two routes R-2 and R-4 which leads to a new route, R-6, consisting of four route
states \{S-2-1, S-2-2, S-2-3, S-4-1\}.

Both R-5 and R-6 could be used as the source for the new route description. However, R-5 has two potential locations for action inconsistencies. The first is the switch from S-1-1 to S-3-1, and the second is the switch from S-3-1 to S-4-1. The alternative route R-6 only has one potential location for an action inconsistency at the switch from S-2-3 to S-4-1. Thus, according to the heuristic, R-6 would be used as the source for the new route description even though it has more route statements than R-5. Since R-6 is based on fewer routes, there is less of a chance that an action inconsistency will occur in the new route description.

The goal of heuristic-2 is to favor the use of complete route descriptions over partial route descriptions. The route descriptions that are used to build the digraph represent complete routes. In a CRISS-based systems, users will have ensured that the route descriptions are safe and coherent and that the routes’ statements are tagged with appropriate landmarks. When the path inference uses these route descriptions to build new route description, heuristic-2 instructs the system to attempt to use complete routes rather than parts of routes, since complete routes represent a complete set of sequential thoughts that have been recognized and approved by the users of the system.

Figure 6.11 demonstrates the effect of the heuristic on the path inference process. The
Figure 6.11. Deciding the path from L-3 to L-429.

The figure represents a digraph consisting of the four routes R-11, R-12, R-13, and R-14. The system has been asked to find a new route description that describes how to get from L-3 to L-429. There are two paths in the digraph that could be used. One potential path passes through landmark L-99. This path starts with the first statement of R-12, S-12-1, and at L-99 switches to R-11’s second statement S-11-2 ending with S-11-3. This route, R-16, would result in a three-statement description, [S-12-1, S-11-2, S-11-3], with one action inconsistency at the route switch at L-99. The second potential path, R-17, passes through landmark L-823 using routes R-13 and R-14 resulting in another three-statement description [S-13-1, S-13-2, S-14-1]. This second route description also has one action inconsistency due to the route switch at L-823.

Both of the potential route descriptions have three route statements, and both have one action consistency. The difference between the two is where the action inconsistency occurs. The potential route description R-17 only uses complete paths. The entire set of
statements from R-13 is used, and the route draws to its natural conclusion at L-823 before being combined with R-14’s statement which also represents a complete set of thoughts. Although there is switch between from one route description to another, it occurs at a logical point where one complete set of thoughts ends and another set begins. On the other hand, the potential route description R-16’s action inconsistency occurs in the middle of a route. It begins at a logical point, the first statement of R-12, but at L-99 switches to the second statement of R-11. The inconsistency of S-12-1 followed by S-11-2 is that S-11-2 was originally written to be preceded by the statement S-11-1. Since the route change occurs in the middle of R-11, the action inconsistency occurs in the middle of a route rather than at the end. While user editing can resolve the issue, heuristic-2 attempts to avoid this situation in the first place. It basically says that it is preferable to join complete thoughts, or complete routes, than to join partial thoughts or routes. Thus, in this situation, R-17 would be the route description returned from the path inference process.

The third heuristic, heuristic-3, states that, everything else being equal, it is better to create new route descriptions using as few route statements as possible. Longer route descriptions are harder to remember than shorter route descriptions. Therefore, in situations where heuristic-1 and heuristic-2 do not apply, heuristic-3 is used to build the new route description from the fewest number of route statements. The goal is to prefer building short route descriptions instead of longer descriptions.

Figure 6.12 shows how heuristic-3 is used. The digraph has been built from four routes, R-23, R-24, R-25, and R-26. The user is requesting a new route description from landmark L-172 to landmark L-311. There are two potential paths. One potential path, R-27, would use R-24 and R-25 resulting in a three statement route description, [S-24-1, S-24-2, S-25-1]. The second potential path, R-28, is built from R-23 and R-26 creating a two-statement description, [S-23-1, S-26-1]. Both R-27 and R-28 are created from two routes so each will have one action inconsistency. Since the number of inconsistencies is the same, heuristic-1 does not apply. Both routes also use the entire set of statements from their source routes, so heuristic-2 does not apply as well. Therefore, R-28 would be the route description returned
to the user since it contains the fewest number of route statements in accordance with heuristic-1.

The three heuristics are applied to the path inference process by associating weights, or costs, with the edges in the digraph. There are four possible weights that can be assigned to an edge.

1. TO\_LANDMARK\_COST - Assigned to an AssociationEdge.

2. TO\_STATEMENT\_COST - Assigned to a PrecedenceEdge that is not associated with the route’s start landmark.

3. START\_ROUTE\_COST - Assigned to a PrecedenceEdge that connects a landmark node to a statement node. The landmark node is that starting landmark for a route, and the statement node is the first route statement in that route.

4. ROUTE\_CHANGE\_COST - Assigned at run-time to edges associated with paths that change routes in the middles of descriptions.

The TO\_LANDMARK\_COST is assigned to AssociationEdges, i.e., edges starting at a statement node and ending at a landmark node. The value of TO\_LANDMARK\_COST
is always 0. This value signifies that the connection does not affect the traveler. Landmark nodes and their AssociationEdges are a computational feature used to indicate that two or more routes share a landmark and may intersect, allowing a path to be created that uses elements from the routes. When following the new route description, the traveler is not affected by the presence of the landmark or lack thereof, so no weight is given to the landmark mention. Edges are assigned these costs by CONNECT(), as shown in Figure 6.7.

The TO_STATEMENT_COST represents the cost of following a route statement. The value of TO_STATEMENT_COST is always 1. It is assumed that routes with more route statements will be more complex than routes with fewer route statements. Therefore, following heuristic-1, longer route descriptions will have higher costs associated with them. These costs are set by CONNECT(), as shown in Figure 6.7.

The START_ROUTE_COST represents the start of a new route, connecting the starting landmark and the route’s first route statement. According to heuristic-1, new paths should be built using as few original routes as possible. To ensure this heuristic is followed, START_ROUTE_COST is set to a high value. The initial PrecedenceEdges at the start of routes have this weight set in CONNECT(), as shown in Figure 6.7. However, there is one special case that may happen when a route is requested. When a route description is requested, the starting landmark of the requested route may not be a starting landmark of any existing route. Therefore at run-time, the weights of the edges leading from the requested route’s starting landmark will be updated to START_ROUTE_COST.

The ROUTE_CHANGE_COST, used to apply heuristic-2, is different from the other costs because it is determined based on the direction through which a node is reached. The ROUTE_CHANGE_COST will be greater than or equal to START_ROUTE_COST. The higher value signifies that while it is possible to join descriptions in the middle, it is preferable not to do so.

Figure 6.13 shows an example of where this cost would need to be calculated. Figure 6.14 shows the same area but with the default weights as set by CONNECT(). The figures show a portion of a digraph where two routes, R-8 and R-11, intersect at the common
Figure 6.13. Example location where ROUTE CHANGE COST must be applied.

Figure 6.14. Digraph with default weights set by CONNECT().
landmark L-41. When the search algorithm is processing S-8-7, it needs to determine the cost to move to the next statement nodes, S-8-8 and S-11-6. Since S-8-8 is part of the same route description as S-8-7, the cost to make that change is TO_STATEMENT_COST, a low cost of only 1. On the other hand, creating a path using S-11-6 is more expensive because it creates an action inconsistency due to the fact that it is a statement node from another route description. Therefore, the edge weight connecting L-41 and S-11-6 is increased from its default weight of TO_STATEMENT_COST to ROUTE_CHANGE_COST.

On the other hand, if the statement node being processed is S-11-5, the cost of moving to the two nodes is opposite that of when starting at S-8-7. Moving from S-11-5 to S-11-6 will maintain the default value TO_STATEMENT_COST since they are nodes in the same route description. But when calculating the cost to move from S-11-5 to S-8-8, the edge weight connecting L-41 and S-8-8 will be increased to ROUTE_CHANGE_COST, since they are statements in different routes.

In order to handle this situation, the digraph is modified immediately before a search. The modification removes the landmark nodes from the digraph and directly connects the statement nodes that precede the landmark node with the statement nodes that originally followed the landmark node. Each new edge receives a weight. If the two statement nodes are part of the same route description, the weight from the original PrecedenceEdge leaving the landmark node is used. If the two statements are from different route descriptions, the new edge’s weight is set to ROUTE_CHANGE_COST.

Figure 6.15 demonstrates how the digraph from Figure 6.14 is modified by this process. In the original graph, S-11-3 is connected to L-89 which is then connected to S-11-4. When the graph is modified, S-11-3 is connected directly to S-11-4. Since the two statement nodes are part of the same route description, the edge weight originally assigned to the PrecedenceEdge leaving L-89 is assigned to the new edge.

Modifying the route intersection at L-41 ends up removing L-41 and creating additional edges, some with weight ROUTE_CHANGE_COST. S-8-7 and S-8-8 are joined and since they are the same route, the weight from the PrecedenceEdge leaving L-89 to S-8-8 is used.
When connecting S-8-7 to S-11-6, the edge weight ROUTE_CHANGE_COST is used to the change in routes. The process for connecting S-11-5 to S-8-8 and S-11-6 is the same.

### 6.5 Path Inference Algorithm

The path inference algorithm consists of three basic steps. First, the digraph is cloned, and the modification described to handle heuristic-2 in the previous step is applied. The modified digraph is then processed to find the shortest path between nodes. If a path is found, the path between the desired starting landmark and the end landmark is used to create a new route description from the route descriptions along the path.

The basic algorithm, INFER PATH(), shown in Figure 6.16, requires three arguments. The graph $G$ is the digraph created by BUILD_GRAPH() from the set of known route descriptions. The startLandmark and endLandmark are the desired starting and ending points for the route for which the user would like a description. It is assumed that the original set of route descriptions has already been searched for a route description that describes a route between startLandmark and endLandmark and that one was not found.
Function INFER_PATH(G, startLandmark, endLandmark)

1. \( mG \leftarrow \text{MODIFY\_DIGRAPH}(G, \text{startLandmark}, \text{endLandmark}) \);
2. \( \pi \leftarrow \text{DIJKSTRA}(mG, \text{startLandmark}) \);
3. \( \text{description} \leftarrow \text{BUILD\_DESCRIPTION}(\pi, \text{startLandmark}, \text{endLandmark}) \);
4. return description;

Figure 6.16. INFER\_PATH() for inferring a new route description for the route from \text{startLandmark} to \text{endLandmark}.

The MODIFY\_DIGRAPH() function, shown in Figure 6.17, is responsible for modifying the original digraph \( G \), as described in Section 6.4. The modification is primarily responsible for ensuring that there is a higher cost for joining route descriptions at the middle route statements, as defined by heuristic-2. In the modified digraph \( mG \) that is returned, there will only be the two landmark nodes \text{startLandmark} and \text{endLandmark}. The \text{startLandmark} node is retained in \( mG \) because it defines the start node for DIJKSTRA(), an implementation of Dijkstra’s algorithm. The \text{endLandmark} node is retained in \( mG \) because it, along with \text{startLandmark}, is used to build the new route description in BUILD\_DESCRIPTION(). The remaining landmark nodes are not retained in \( mG \), but are instead transformed into weighted edges that determine if statement nodes are part of the same route or different routes.

The edges from the original digraph are checked, and if they start or end with the two special landmarks (lines 8 and 12 respectively), or connect two statements (line 10), the original edge and its original cost are used in \( mG \). Otherwise, at line 14, the end node of the edge is a non special landmark. The nodes to which it is connected and their route ids are check to see if they are different from the current vertex \( v \). If they are the same route id, signifying two consecutive route statements in the same route descriptions, the new edge uses the original cost of the edge from the landmark node to the second statement node. If two statement nodes come from different route descriptions, the new edge is assigned the higher weight ROUTE\_CHANGE\_COST.

The modified digraph \( mG \) returned from MODIFY\_DIGRAPH() is then passed to Dijkstra’s algorithm [14] (Figure 6.18). Using \text{startLandmark} as the initial node, the
Figure 6.17. MODIFY_DIGRAPH() for removing landmark nodes and adjusting edge weights.

algorithm finds the least cost path from startLandmark to all nodes in the graph. During processing, π is used to maintain the pairs of nodes used to build the optimal path from startLandmark to all nodes. At the end of processing, π is returned so that the new route description can be built.

The last step in the process, BUILD_DESCRIPTION() shown in Figure 6.19, builds a route description using the least cost path found by Dijkstra’s algorithm. The new descrip-
Function DIJKSTRA($mG$, $startLandmark$)

1. $\text{foreach vertex } v \in V[mG] \text{ do}$
2. \hspace{1em} $d[v] \leftarrow \infty$;
3. \hspace{1em} $\pi[v] \leftarrow \text{NIL}$;
4. $\text{end}$
5. $d[startLandmark] \leftarrow 0$;
6. $S \leftarrow \emptyset$;
7. $Q \leftarrow V[mG]$;
8. $\text{while } Q \neq \emptyset \text{ do}$
9. \hspace{1em} $u \leftarrow \text{EXTRACT\_V\_WITH\_MIN\_D}(Q)$;
10. \hspace{1em} $S \leftarrow S \cup \{u\}$;
11. \hspace{1em} $\text{foreach vertex } v \in \text{Adj}[u] \text{ do}$
12. \hspace{2em} $\text{if } d[v] > d[u] + w(u, v) \text{ then}$
13. \hspace{3em} $d[v] \leftarrow d[u] + w(u, v)$;
14. \hspace{3em} $\pi[v] \leftarrow u$;
15. $\text{end}$
16. $\text{end}$
17. $\text{return } \pi$;
18. $\text{return } \pi$;

Figure 6.18. Dijkstra’s algorithm for solving single-source shortest path.

The original route descriptions are natural language texts, they lack structure that a computer can easily exploit. This limits the amount of direct processing that can be done with the data set. RAE and CRISS, through the addition of landmark tags on
Figure 6.19. BUILD_DESCRIPTION() for building a new route description.

individual route statements, provide an element of structure to the text. This additional information, though limited, allows the descriptions to be processed by simple, well-known algorithms after transformations.

The path inference process is both an example of how the CRISS/RAE structure can be used for additional processing as well as for solving a real problem faced by people with visual impairments. Assuming that the set of route descriptions contains routes that intersect in the real world and are tagged appropriately, the algorithm can find real-world routes without a deep understanding of the descriptions. Following information extraction, which is based on a shallow understanding of text, the path inference process is based on a shallow understanding of route descriptions. The guiding principle is that when natural language route description are tagged appropriately with a set of landmarks, landmarks shared by different routes can be used as markers to combine parts of the different routes into a new description.
There are limitations to the algorithm. If there are insufficient numbers of route descriptions sharing common landmark tags, the system may be unable to find a path from a starting landmark to a destination landmark. If the set of route descriptions is sparse and only a few descriptions have been entered into the system, the system may have a difficult time finding routes. If the set of route descriptions is sparse but still contains a routes sharing landmark tags, the system may be able to find a route description between landmarks, but the description may describe a route that meanders, is too long, or is far off the path a person would normally travel. A low number of route descriptions in the system will prevent the system from performing at a level that is useful for the users.

In order for this algorithm to be successful, it is necessary that the set of route descriptions is growing and is constantly managed by the CRISS-user population. Assuming that a CRISS-based website has an active community of users, this should not be a problem. It is should also be assumed that any new route description generated by this process will be checked by knowledgeable users in order to verify its safety, accuracy, and usefulness.

Another limitation with the current design is that it does not handle path reversal. Path reversal is the situation in which there is a path from landmark A to landmark B and the reverse path from landmark B to landmark A. When walking, a person can go from A to B, turn around, and go from B to A. Ideally, this situation should be handled by RAE. A route description could be entered by a user describing how to get from A to B. Instead of requiring the user to enter another route describing how to get from B to A, the system would automatically generate a route in the reverse direction. As described in this work, RAE cannot currently produce the reverse path description.

RAE cannot handle path reversal for two reasons. The first reason is due to the graph used by the path inference process. In the graph, the precedence relationship that places one statement before another statement in a description is represented as a directed edge. The reason for this design decision is that many blind individuals, particularly those with complete vision loss, trail walls. Since there may be one set of landmarks mentioned when following one wall and another set of landmarks mentioned when following the opposite
wall and going in the return direction, a directed edge was chosen to handle this situation. Since there are paths where the reverse path may not have any different landmarks than the forward path, an adjustment in the representation may be made to the representation in order to handle the reverse path.

The other reason for lack of path reversal support is that it is more difficult to address. A route description is a set of route statements, and a route statements is currently defined as a single sentence. A single sentence may describe one or more actions that need to be performed along the route. When a new route description is built, the route statements are collected in order without modification. If the order of the route statements were simply reversed, the result description would not necessarily describe the path in reverse order. For example, a route description describing how to go from A to D may look something like, “From A go to B, and then from B go to C. At C, go to D.” which implies $A \rightarrow B \rightarrow C \rightarrow D$. Simply reversing the two route statements - “At C, go to D. From A got B, and then from B go to C.” - would imply the landmark sequence to be $C \rightarrow D \rightarrow A \rightarrow B \rightarrow C$ which is not the original path in reverse. To handle this problem, the sentences, and even individual phrases representing a single action, may have to modified to represent the reversed route description.

Although the three heuristics are intended to reduce the occurrences of action inconsistencies, inconsistencies can still be present in the generated route descriptions and remain a system limitation. A common case is when a turn is required at the point in a new route description where two known route descriptions have been joined. For example, in the new route direction created in Section 6.4, an extra statement had to be added by a knowledgeable user in order to identify the turn in travel that needs to occur at the Quad’s sidewalk intersection. Currently, the system makes no attempt to identify these situations or correct them.

In this iteration of RAE, a commitment was made to shallow language understanding. The goal was a system design capable of processing natural language text and relatively simple in design and algorithms. In order to address these two limitations, an additional
understanding of route direction could be added to the system. Although this may increase the complexity of the system, the additional functionality and complexity may result in improved route generation. Potential enhancements are addressed in Chapter 7, Future Work.
CHAPTER 7

FUTURE WORK

The work presented on ShopTalk and RAE proposes alternative approaches to assistive navigation technology for independent blind travelers. Both systems still have unexplored questions and other avenues for research. This chapter details future work and research for each system.

7.1 ShopTalk

The ShopTalk experiments discussed in Chapter 3 were necessarily limited in scope. They were limited to shopping for shelved products while carrying a shopping basket. In order to fully realize an assistive shopping solution for the visually impaired, additional issues need to be addressed.

The first obvious area of work is the hardware. There are several reasons why the hardware should be addressed. First, ShopTalk is primarily a software solution. In order for a ShopTalk-like system to be widely adopted, it should be capable of running on common, commercial, and relatively cheap hardware. In order for ShopTalk to be adopted, the stores and shoppers must be able to afford the device. Second, the current hardware is a large and bulky design that is not ergonomic. Smaller hardware would be easier to store between uses, have lower power requirements, and be easier to carry while shopping. Although the participants in these experiments were not asked how they felt wearing the backpack, other navigation experiments [48, 54] have noted that their participants mention they feel conspicuous when using such a device. More compact hardware would be less noticeable to other shoppers and help users avoid feeling that they are overly conspicuous when using the device.

With these issues in mind, ShopTalk is being ported to run on smaller and more
manageable hardware. Figure 7.1 compares the hardware used in the experiments reported in this work with the hardware for the next version. The new version is being designed to run on a Nokia smartphone with a Bluetooth barcode scanner and a Bluetooth ear piece. The new barcode scanner is a different type of scanner, a pen-based scanner as opposed to the 2D scanner used in the experiments. The new set of hardware has no cables, no backpack, and uses a common cell phone.

Another limitation of the reported experiments was that the device was only used to locate products on shelves in aisles. In order for ShopTalk to become a more full featured tool that can address the full shopping experience, other areas of the grocery store need to be handled by the system. Other store areas to be mapped include:

1. Refrigerator and freezer units requiring shoppers to open a door.

2. Refrigerator units with no doors. These units are similar to shelves except they have bottom areas that hold the cooling units and extend out from the shelves. Another
style are the large, refrigerated boxes that shoppers reach down into from an open
top.

3. Deli and bakery areas that may or may not have people working behind a counter
and food behind glass barriers.

4. Produce areas, especially those areas that have free-standing carts that hold produce.

Each type of area presents a different set of challenges. Produce areas, in particular,
will likely be problematic. Produce often requires a visual inspection to ensure that the
food is of good quality. It may simply be the case that ShopTalk is unable to guide all
shoppers to all areas and all products in a grocery store. Yet, if the system can be used to
find the majority of items in a store, it will still be useful to people with visual impairments.

A third limitation of the experiments was that the participants only carried shopping
baskets. For shopping trips in which a shopper needs to buy heavy items or a large number
of products, a hand basket would be insufficient. Therefore, it would be useful to test how
well visually-impaired shoppers can navigate a grocery store when using a wheeled shopping
cart. The greatest concern would be collisions with other shoppers, shelves, and displays.
During conversations with the participants of this paper's experiments, several noted that
they use shopping carts for their regular shopping. They stated that rather than pushing
the cart, they walk in front of the cart and pull it behind them. So, for at least part of the
visually impaired population, using a shopping cart with ShopTalk would appear possible.

Due to scope limitations, this work does not directly address or test product identifi-
cation. The experiments presented in this work were concerned with guiding the shopper
to the exact location of the shelf barcode beneath the target product. Once the shoppers
scanned the correct barcode, they were instructed to pick up the item directly above the
barcode. The issue with that is that it is not always correct to assume that products are
in the correct location. Another problem is that a shopper who is visually impaired may
deviate slightly right or left and pick up an incorrect product. In order to alleviate this
problem, a second stage of product identification could be added. This would ensure that
the product the shopper picked up was the intended product. As mentioned earlier, for some people with partial vision, the task of product identification could be accomplished by visual inspection of the item. For the people unable to visually verify the item, it would be possible to either perform a second barcode scan of the UPC barcode or integrate one of the technologies from the related work section such as the kReader Mobile [57] for OCRing images of the product.

A final limitation of this work is that it does not address how to build the shopping list. There are two types of shopping, planned shopping and spontaneous shopping. In planned shopping, a person makes a list of all items that need to be bought. In spontaneous shopping, the shopper is already at the store and decides that a item that is not on the planned shopping list is needed. Adding this functionality to ShopTalk would complete the shopping experience and ensure that ShopTalk allowed people with visual impairments to perform the entire shopping experience independently.

7.2 CRISS and RAE

The work on CRISS and RAE, up to this point, has not involved research that integrates the target users of CRISS. The work described herein targeted RAE and its functionality. The CRISS framework is introduced, but its purpose and functionality has only been outlined. Consequently, there is opportunity for work focusing on CRISS and for integrating RAE into that system.

The work described here contains only a limited number of route descriptions, a direct consequence being that the JAPE rules used for autotagging had to be built manually. The problem with this method is that it is error prone and time consuming. One way to address this problem would be to collect a larger set of route descriptions. A larger set of route descriptions would enable the use of machine learning techniques to automatically build the rules. This would ensure that as more route descriptions are added to the system, new rules could be added quickly and existing rules would be easily adapted or replaced.

Another benefit of a larger collection of route descriptions is that more extensive route analysis would be possible. It is likely that people use language and description patterns
other than those identified during analysis of this work’s route description set. An additional benefit of this type of analysis is that existing systems that generate route directions could be analyzed as well. By comparing their output to description styles and language patterns that people with visual impairments generate, it may be possible to improve the quality of the automatically generated route descriptions leading to descriptions that are better suited for blind travelers.

The set of route description that were analyzed for RAE were originally submitted by participants from various regions of the United States. The advantage of this approach is that it led to descriptions of routes and environments that do not exist at USU. The disadvantage of the approach is that it makes it difficult for comparison and testing purposes. It was not possible, for example, to actually test someone attempting to follow one of the routes described. Therefore, in addition to increasing the overall set of route descriptions as mentioned above, another set of route descriptions should be collected that is specific to a known environment, such as the university campus. This would enable analysis and experimentation of the relationships between the route descriptions and execution of the descriptions. With this type of collection, descriptions from multiple participants for the same route could be collected, and differences and similarities between the routes could be analyzed.

Another area for work is to build a wiki-style website based on the CRISS framework. This would involve designing and building the website and testing the system with the target users. Once the website is built, it will be possible to examine how people take advantage of the ability to create and edit route descriptions. The CRISS framework is based on wiki-style functionality, allowing users to upload and edit routes. In a traditional wiki, the list of changes made to a document are maintained so that people have the ability to examine the history of a document. Maintaining and then analyzing the edit history of route descriptions will help to gain insight into which features of route descriptions people consider useful and which features people deem less helpful or confusing.

Another avenue of research would be to have sighted people add route descriptions
to the system. Many university centers devoted to helping students with disabilities have
to build up a initial set of route descriptions covering the campus. This would serve two purposes. First, it would help build the campus map at a quicker pace. Another purpose would be to track the changes individuals with visual impairments made to route descriptions written by sighted individuals. Tracking these changes would help build a better understanding of the differences between route descriptions written by blind and sited travelers. This could help create training guideline for sighted users on how they should write routes descriptions that are meant for blind travelers.

Tracking user edits in CRISS would also allow the differences in the route descriptions created by users with different profiles to be compared and analyzed. When describing vision loss, terms such as “blind” and “visually impaired” are often used, yet there are many variations of the disability. Some people have total vision loss, and others have varying amounts of residual vision. The length of time that people have had vision loss also varies. It may be that these variations in the amount and type of vision loss affect how people write route descriptions. By associating profiles with users that help to identify a person’s type of vision loss, and then analyzing the changes made to their route descriptions and the changes they make to others’ route descriptions, it may be possible to identify language patterns that are particular to one group or another.

Once CRISS is built, RAE can be integrated into the system. This would create an opportunity to investigate how users would actually use the landmark tagging system. Once users have tagged a number of route descriptions, this could be used a create a list of the type of landmarks that visually-impaired travelers are truly interested in. This would help inform other map builders what features should be included on maps targeting blind users. It would also help improve the autotagging features in RAE in that only the most useful landmarks would need to be targeted by the JAPE rules, thereby increasing the effectiveness of the system.

As mentioned in Chapter 6, the path inference process has two limitations, path reversal
and action inconsistencies. These may be dealt with in such a way that the quality of generated route directions are improved while still staying within the RAE's commitment to shallow language understanding. The number of action inconsistencies may possibly be reduced by analyzing certain keywords that occur when different route directions are combined in new route descriptions. For example, cardinal directions are used often in route directions given by the visually impaired. It may be possible to identify a potential action inconsistency if “North” is mentioned in one sentence from one original route description and then “East” is mentioned in the next sentence from another route description. Identifying such a situation could signal the system to insert an additional, system-generated instruction such as “Turn right” at the location in the generated route description where the original route descriptions are connected. Using keywords, such as cardinal directions, as identifiers, it may be possible to reduce the number of action inconsistencies created when new routes are built.

Another avenue of investigation for developing methods to reduce action inconsistencies is to incorporate GIS-based maps into the system. It may be possible to map landmarks mentioned in route directions to location on maps. It may also be possible to develop a method that compares the generated natural language route description to the route on the map. If the generated text describes that passes through decision points such as turns, the system could add additional instructions to the text.

Addressing the path reversal limitation will require more work. As mentioned earlier, it is not simply a matter of reversing the order of the route statements or sentences. The order of words within a sentence identify a specific set of actions. Using additional information extraction techniques, one may be able to break sentences into individual action phrases allowing phrase to be reversed. This process will not eliminate the issue, however, since even simple phrases, e.g., “from A, go to B,” require that the wording of the phrase must be altered in order to indicate that one has to go from B to A on the reverse path. Work will need to address the task of not only reversing the actions, but reversing the actions described in the text so that the landmarks are experienced in reverse order using
appropriate language.

This work has been focused on guiding people with visual impairments using only verbal route descriptions. Another avenue for future work is to relax that restriction and investigate integrating verbal route descriptions with existing ETAs. For example, in outdoor settings, landmarks could optionally be associated with a latitude and longitude. It would then become possible to experiment with and extend GPS-focused ETAs such as Sendero GPS. The question then becomes do route descriptions written by experienced visually impaired travelers improve the user experience with GPS ETAs? Integration would also allow for extending outdoor-based systems to include routes in indoor environments. Outdoor localization would be handled by the standard GPS-based ETA and indoor environments would be handled using only verbal route descriptions.
CHAPTER 8
CONCLUSION

Many research projects have designed assistive navigation devices for the visually impaired. Other projects have designed assistive navigation devices for the general population that could be adapted to use by people with visual impairments. Unfortunately, only a small number of these have made the transition from research-oriented projects to commercial products. In the commercial arena, only a few products based on GPS have had any measurable success in the market as ETAs for the visually impaired, but even then the adoption rate remains low among the target population. Navigation support for indoor environments is nonexistent on those devices due to GPS’s technical limitations. The adoption rate of non-GPS navigation products remains low to non-existent due to the cost of infrastructure modification, power issues, and maintenance.

One solution that could help to avoid some of the reasons for the ETA low adoption rates, would be to start treating the person using the device as an integral part of the system rather than just the user. Independent travelers develop a wide variety of navigation skills through O&M training and having to perform daily navigation tasks. Rather than waiting for readings from the device’s sensor set to determine a user’s location, an ETA can assume the user has the appropriate set of skills to follow a set of verbal instructions. This is analogous to the separation of the planning layer and the control layer in robotics and in the SSH [60]: the ETA is the planner layer and the traveler is the control and sensor layer.

This dissertation addresses four hypotheses using two systems as experimental platforms. The first hypothesis was that a navigation assistance system for the blind can leverage the skills and abilities of the visually impaired, and does not necessarily need complex sensors embedded in the environment to succeed. ShopTalk provides an example of such a system. It provides route descriptions for the locomotor space of a real-world grocery
store, and multiple participants were able to follow these descriptions to successfully find their way around the store. In the locomotor space, no additional sensor or device was installed in the environments, only the skills of the participant were used when following the descriptions.

The second hypothesis was that verbal route descriptions are adequate for guiding a person with visual impairments when shopping in a supermarket for products on shelves located in aisles. ShopTalk shows that, at least for part of the visually-impaired population, this hypothesis holds true. In two experiments, participants were guided to individual products located among thousands of other products. The advantage of this type of navigation method is that it shows that, at least in structured environments similar to a sequential aisles in a grocery store, simpler navigation devices can be sufficient and that additional complex sensors are not always necessary. The grocery store is a structured environment that has a specific type of layout that people are familiar with, even if they have impaired vision. Taking advantage of this familiarity and the travelers’ skills helps people with little to no vision find individual products out of thousands. Essentially, a device such as ShopTalk can help solve a real-world needle-in-the-haystack problem.

ShopTalk’s approach worked in the experiments because the area of the store to which the experiments were restricted is very structured. Decision points, such as entrances to aisle, are easy to identify. Aisles are easy to follow. Cashiers have distinct sounds that identify their location. However, this approach was not tested in the whole store, and it remains to be seen if this approach would work in areas of the store such as the produce section. In the produce section of Lee’s Marketplace, the store in which the experiments were performed, there is an open refrigerator section acting as a boundary to the produce section. In the section, rather than shelves, there are large carts with wheels. The wheeled carts make it easy for the section to be reconfigured, adding an extra challenge to a blind navigator. A larger store, such as a Sam’s Club, which contains large amounts of areas without shelves would present extra challenges. It remains to be discovered what environmental configurations may prevent the use of ShopTalk.
The third hypothesis of the dissertation was that information extraction techniques can be used to automatically extract landmarks from natural language route descriptions. RAE was used to provide evidence that this was possible. By analyzing multiple route directions provided by people with visual impairments, patterns in the language were found that were used to develop rules for extracting landmarks. From a set of 52 indoor and 52 outdoor example route descriptions, information extraction was shown to be capable of finding and annotating landmarks once a set of common language patterns that people use when describing a route was identified. In taking advantage of these patterns, we were able to automatically identify landmarks in the descriptions. It is hypothesized that with a larger corpus of route descriptions, additional language patterns might be discovered and the landmark identification process may be improved.

The final hypothesis was that new, natural language route descriptions can be inferred from a set of landmarks and a set of natural language route descriptions whose statements have been tagged with landmarks from the landmark set. The INFER PATH() heuristic demonstrates that the additional structure that landmark tags adds to route descriptions and their statements can be exploited to build new route descriptions. The advantage of this approach is that it only requires a shallow understanding of the language used in route directions, reducing the need for more complex natural language processing. If the new routes contain action inconsistencies, these will be corrected when users review the new descriptions. RAE is intended to be part of a larger system in which users’ familiarities with blind navigation techniques and the target environment will be used as aids for correcting any potential problems with the instructions. The main purpose of this feature is to give users insight into the spatial relationships that exist between the routes with which they are already familiar.

Verbal route directions are a viable option for an ETA. However, one of the advantages of this approach is that devices based on verbal route directions do not necessarily have to be implemented solely as a route instruction-based system. Verbal route guidance can be used to extend the areas for which an existing system is capable of guiding people.
For example, adding a verbal route instructions component to an existing GPS-based ETA would enable the device to give verbal route instructions for indoor environments, areas that are not supported due to GPS’s technical limitations. Even areas covered by the existing devices could be supplemented by verbal route directions. For example, many GPS devices primarily cover street-based routes due to the data source of the maps they use. Verbal route descriptions could supplement outdoor environments such as college campuses where there are large areas of sidewalks and few streets.

Extending existing devices with verbal route descriptions would not drastically increase their costs. This type of system does not require modifications to the environment. Since no device is installed in the environment, there are no power or maintenance issues. Routes would have to be written and monitored to ensure their safety and currency. A CRISS and RAE-based system, following the wiki-style system of sites such as Wikipedia, would ensure that the routes are written by knowledgeable users. The result would be wider route coverage at lower costs.

It is mentioned several times throughout the discussion on RAE and CRISS, that the safety of the routes in the system are a concern. RAE and CRISS are capable only of shallow language understanding and have no connection to a GIS system. The goal is to provide additional navigation information that a user can use in conjunction with their everyday navigation skills. This is the same approach taken by commercial products such as Sendero GPS. In the Sendero GPS manual [108], it explicitly states, “The individual user is wholly responsible for all issues related to personal safety and mobility.” Although the device can compute and present a route for an outdoor environment, the user must remain aware that the device is not a complete replacement for everyday navigation skills, but rather an additional tool. The idea is that, from an independence point of view, it is better to have some information rather than no information at all. Like the commercial products, RAE aims to increase the amount of spatial information available to users and, in the end, increase their navigation independence.
REFERENCES


[17] Crandall, W., Bentzen, B. L., Myers, L., and Brabyn, J. New orientation and accessibility option for persons with visual impairments: Transportation applications for
remote infrared audible signage. *Clinical and Experimental Optometry* 84, 3 (May 2001), 120 – 131.


APPENDICES
Appendix A

ROUTE DIRECTION TRANSCRIPT

The following is the complete set of route directions given by a visually-impaired student at USU. The student produced the directions verbally from memory. The route starts at the Taggart Student Center, goes to the Old Main Building, and finally goes indoors to the elevators in Old Main.

Leaving the DRC, once you’re outside the Disabilities Resource Center’s doors, you are facing southwest. The best thing to do is to walk forward until you reach a flight of stairs that will actually come at you from an angle. Your dog or your cane will find these and you’ll need to make a right turn parallel to those, to these stairs. I believe the easiest thing to do is follow these stairs until you reach an opening on your left. This is a wheelchair ramp. On the left of this wheelchair ramp there is a handrail which pokes, that hangs out into the sidewalk further than the stairs that you’re following. So you can find this. Or you can make a moving turn to your left to enter the wheelchair ramp. One thing you can note if you’ve gone too far is on the west side of the wheelchair ramp there is a large cement structure. This is on the west side of the wheelchair ramp. And on the east side of the wheelchair ramp there is the handrails. So find either of those structures, and turn left which is facing south and begin to walk up the wheelchair ramp. The wheelchair ramp is fairly long. Follow that up until you reach the landing which is an intersection of a sidewalk. Once you reach that sidewalk make a right turn and begin heading west. At this point you will hear, if it’s during the warm season, there will be a fountain coming up on your left. And on your right there is a patio. So there will be things, you can hear cues. People socializing. Often, there’s the smell of cigarette smoke here.
Good cue to keep moving forward until you pass these sounds. The fountain will be now towards your back on your left. And the sounds from the patio will be towards your right on your back. And you will need to make a gradual left turn. This is a large open sidewalk area. The best thing to find is a cement, a large curb about, almost knee high on an average person. This is on your left. So you’ve made your gradual left turn. And hopefully you will find a knee high planter or curbing around a grassy area. If you reach it directly in front of you, if you hit it perpendicularly, then you have made too sharp of a left turn, and you can move to your right and follow this curbing around, and on moving to your right, the curbing will make a 90 degree turn to the west. Follow that and you will feel that it will be pointing you towards a southwest direction again. It will make another 90 degree turn towards the south. And this is the sidewalk which will, again, point you in the direction towards Old Main. If you are on this curb there is a trashcan near this curb after you are facing south if you’ve made a left turn at the planter box. So now we’re facing south, heading on a sidewalk towards Old Main. Follow this sidewalk and it will begin to veer to the right. Stay on this sidewalk as it turns to a southwest direction. You, chance by chance, hear other people walking in a perpendicular fashion. There is an intersecting sidewalk here. You can keep going through that intersection. Still heading southwest. You’ll find that you will hear an exhaust system for this Eccles Science building. Possibly air conditioning or heating on your right; it’s quite loud. Heading southwest on this sidewalk you will hear that exhaust system and you will need to make a moving turn to your left. Now you are facing north. This intersection, right here in the, where you are doing this moving left, it is essentially a four way intersection. There is the sidewalk, that you are following southwest, continues. There is also a sidewalk intersecting directly west, heading east - west. That would intersect the sidewalk you are on. And also there is a sidewalk heading right up to north and that is the sidewalk you
need to be on to get to the main entrance of the Old Main building. So, if you, once you found the intersecting sidewalk that heads directly north.

So, heading south towards the entrance to Old Main. You will most definitely sense, feel, see large open space to your left, this is the Quad, a large grassy area. On your right will be large trees providing shade or the wind could be blowing through the leaves. These trees are around the east side of Old Main. So, heading south you will either trailing the Quad side of the sidewalk or the Old Main side of the sidewalk. On the Quad side of the sidewalk, you will find that there is an intersecting sidewalk, probably - in steps I’m not sure, a lot of steps - in yards, probably 15 to 20 yards. There will be an intersecting sidewalk going off to your left. This sidewalk heads across the Quad, but is a good cue to know to move to your right. Do a moving turn into some ... the essentially kind of ... almost patio stuff before the - I know I’m giving bad terms. So if you find that intersecting sidewalk that goes east across the Quad you can do a moving turn to your right. In this area there is a new surface. I believe you could consider it maybe cobblestone, or a rougher surface. There’s also trees inside planter boxes, although they’re not, there’s not any curbing or anything around these planters so you could possibly run into dirt. Move around these, so you are on the west side of these tree planted areas on to, again, the rough kind of surface, cobblestone. If you’re a guide dog, at this time you could give your dog the instruction to find the stairs if he or she is associated with that. You keep moving to your right which will be the west. You might directly encounter the stairs. Or there will be a cement curbing, maybe knee high, somewhere in the shin area. The stairs are either going to be to your right or to your left. On each side, excuse me, on the north side of the stairs along that curbing there is a waste bin, a garbage can. So if you encounter the garbage can, the stairs are going to be to the south of this garbage only by a few feet. The stairs are wide and there’s a hand rail in the middle of the stairways, kind of dividing the
stairway to right and to left. On each side of the stairs there are hand rails as well and a cement structure trailing up the stairs. So if you reach the bottom of the stairs, facing the stairs, you are now facing west. Ascending the stairs you’ll reach the landing which is five steps before the main doors to the Old Main building.

So once you found the front doors to the Old Main building, you will enter them. There’s multiple doors and it seems some doors swing both directions. So once you’ve entered into the Old Main, there’s a rug. After the rug there’s a tiled surface. This is the entrance to the Old Main, would that be called the lobby in there? So this is the lobby of the Old Main building. On your right there is some computers so you might be able to hear some people using the keyboard. But just keep heading deeper into the building there’s another set of doors. Again there’s multiple doors to choose from, and I believe doors swinging in both directions. So once you enter into, enter through those doors, there will be carpeted surface and this is a main hallway heading east-west inside Old Main. There’s classrooms on both sides. Move across this room, or through this hallway and you will find that the hallway, this is a large hallway and to get to the rest of this hallway, the hallway narrows on the left side. So make a gradual turn, and if you’re using a guide dog, you can make a moving turn left and your dog will find the hallway. If you’re using a cane there, you could run into a wall that tapers the large hallway into the small hallway. So now that you’ve entered the small hallway you are still heading west inside Old Main. You will come upon an intersecting hallway. You will hear that the hallway runs north-south across the length of Old Main. Once you reach this intersecting hallway, you need to make a left turn. Now you’re heading south. Make approximately 5 to 10 steps, you will most likely be able to hear a hallway on your left which harbors the elevators. Make a left turn into this small hall or corridor and the elevators will be on your left, on the north side. The elevators are facing south.
There’s two elevators. The up/down buttons for the elevator are right in the middle of both of these elevators.
Appendix B

NAMED ENTITY SETUP ANNOTATIONS

The following describes the words and phrases annotated during the setup phase of RAE’s autotagging named entity recognition.

B.1 Verbs

Annotates the forms of “to be” and “to have.”

- Forms of “to be”
  - “is,” “are,” and their contractions “’s” and “’re.”
  - Phrases consisting of modal verbs preceding “be.”

- Forms of “to have”
  - “have,” “has,” and the contraction “’ve.”
  - Phrases consisting of modal verbs preceding “have.”

Modal verbs are identified by ANNIE’s POS tagger [20]. They are verbs that do not take an “s” ending in the third person singular present. For example, “can,” “could” and “should.”

B.2 Cardinal Directions

Annotates forms of cardinal directions.

- Abbreviations
  
  N, S, E, W, N., S., E., W., NE, NW, SE, SW

- Words
north, north-, northern, northeast, northeast-, northeastern, northward, northwest, northwest-, northwestern south, south-, southern, southeast, southeast-, southeastern, southward, southwest, southwest-, southwestern east, east-, eastern, eastward west, west-, western, westward

Phrases consisting of multiple directions are annotated as well, e.g., “north northeast.”

B.3 Distance
Annotates terms and phrases related to distance.

- Measurement terms
  foot, feet, ft, inch, inches, meter, meters, metre, metres, mile, miles, paces, step, steps, yard, yards

- General terms
  distance, length, width

The measurement terms may be preceded by:

- Numbers as in “5 feet” or “10 or 12 inches.”
- The phrase “and a half,” as in “six and a half meters”

The general terms are preceded by a determinant and optional adjectives identified by ANNIE’s POS tagger, as in “the width” and “a short distance.”

B.4 Simple Transitive
Annotates simple transitive prepositions. These were identified in Jackendoff’s language analysis [53].

about, above, across, after, against, along, alongside, amid, amidst, among, amongst, around, at, atop, before, behind, below, beneath, beside, between,
B.5 Intransitive

Annotates intransitive propositions. These were identified in Jackendoff’s language analysis [53].

afterward, afterwards, apart, away, backward, backwards, downstairs, downwards, forward, forwards, here, inward, inwards, left, outward, outwards, right, sideways, there, together, upstairs, upward, upwards, straight, straight ahead

Jackendoff’s analysis originally covers N-wards words including terms such as “northwards” and “southwards.” Since these are already covered in the cardinal direction rules, they are omitted here.

B.6 Compound Transitive

Annotates compound transitive phrases. These were identified in Jackendoff’s language analysis [53].

in back of, in between, in front of, in line with, on top of, to the left of, to the right of, to the side of

B.7 Angle

Annotates short phrases referring to angles. The terms are

angle, angles, degree, degrees

These terms must be preceded by one of the following:

• A number, as in “45 degrees.”

• The words “left” or “right,” as in “left angle.”
B.8 Biased Part

Annotates words and phrases that refer to parts of landmarks. Terms for parts are back, beginning, bottom, center, edge, end, ends, front, level, middle, rest, side, sides, top

The terms can optionally be modified a several ways:

- A cardinal direction, as in “north end.”
- Directional adverbs identified by ANNIE’s POS tagger. This overlaps with the cardinal directions.
- A determinant, as in “the edge.”
- Adjectives identified by ANNIE’s POS tagger, as in “low side.”
- The specific modifiers:
  far, near, righthand, lefthand, left, right, left hand, right hand

The terms and modifiers are then used to identify three types of biased part phrases:

- A spatial intransitive, followed by an optional modifier, followed by one or more part terms, e.g., “at the end” and “on the far right hand side.”
- A required determinant followed by an optional modifier, followed by one or more part terms, e.g., “the middle” and “the main level.”
- A coordinating conjunction, followed by an optional modifier, followed by one or more part terms, followed by a preposition, e.g., “and edge of.”

B.9 Egocentric Reference

Annotates words referring to the traveler himself.

- Self references:
you, yourself

- Reference to part of self. The term “your,” followed by an optional set of adjectives, followed either one or more nouns or one of these specific terms:
  
  back, body, feet, foot, hand, hands, left, path, right, self, side, way

- Guide tools:
  
  your cane, your dog, white cane, guide dog, your white cane, your guide dog
Appendix C

EXAMPLE PATH INFERENCES

The following are examples of routes processed by the path inference process. In all examples, it can be assumed that the source routes have all been tagged and approved by knowledgeable users. If a landmark is used as a tag, it is guaranteed to be located in the landmark hierarchy. It is also assumed that the inferred route descriptions were the built from the lowest cost paths in the digraph.

C.1 Example 1

This example demonstrates the simple concatenation of two source route descriptions. The first source descriptions describes a route from the Animal Science Building to the east entrance to Old Main on the USU campus.

ROUTE_ID: R-81
START LANDMARK: Animal Science
END LANDMARK: Old Main East Entrance
STATEMENTS:

1. Exit Animal Science at the south entrance. [Tag: Animal Science]
2. Walk forward to you detect the entrance to the quad. [Tag: Quad]
3. Keep walking south until you detect the sidewalk intersection in the center of the quad. [Tag: Center of the Quad]
4. Turn right, walking west.
5. Keep walking straight until you detect bricks on the ground.
6. Walk straight until you detect a wall.
7. Turn left, and look for stairs on your right.
8. Go up the stairs to the entrance. [Tag: Old Main East Entrance]

The second source route description describes the route from the entrance to Old Main to the entrance to the President’s Office which is located inside Old Main.

**ROUTE ID:** R-89

**START LANDMARK:** Old Main East Entrance

**END LANDMARK:** President’s Office

**STATEMENTS:**

1. Enter the building from the east entrance. [Tag: Old Main East Entrance]
2. Walk forward until you find a second set of doors.
3. Go through them and walk down the long hall.
4. Keep walking straight.
5. You will eventually run in a wall at a t intersection.
6. You should be able to detect some glass doors.
7. These are the doors to the President’s Office. [Tag: President’s Office]

At this point, a user asks the system to find a route description for a route from Animal Science to the President’s Office. Since R-81 ends at the landmark “Old Main East Entrance” and R-89 begins at the same landmark, the path inference process would concatenate these two routes into a new route description. The resulting route would be:

**ROUTE ID:** Waiting for user approval

**START LANDMARK:** Animal Science

**END LANDMARK:** President’s Office

**STATEMENTS:**

1. Exit Animal Science at the south entrance. [Tag: Animal Science]
2. Walk forward to you detect the entrance to the quad. [Tag: Quad]
3. Keep walking south until you detect the sidewalk intersection in the center of the quad. [Tag: Center of the Quad]

4. Turn right, walking west.

5. Keep walking straight until you detect bricks on the ground.

6. Walk straight until you detect a wall.

7. Turn left, and look for stairs on your right.

8. Go up the stairs to the entrance. [Tag: Old Main East Entrance]

9. Enter the building from the east entrance. [Tag: Old Main East Entrance]

10. Walk forward until you find a second set of doors.

11. Go through them and walk down the long hall.

12. Keep walking straight.

13. You will eventually run in a wall at a t intersection.

14. You should be able to detect some glass doors.

15. These are the doors to the President’s Office. [Tag: President’s Office]

This route description does not contain a noticeable action inconsistency. The statements where the two source routes are joined, statements 8 and 9, are a logical set of actions. Therefore, this description would require little, if any, editing from the user during the approval process.

C.2 Example 2

This example is similar to the previous example, but it demonstrates how action inconsistencies can arise when combining route descriptions. As with the previous example, there are two source route descriptions. The first description describes the route from Room 405, also known as the CSATL Lab, to the Room 414, also known as the Computer Science front office.
ROUTE_ID: R-118
START LANDMARK: Room 405
END LANDMARK: Room 414
STATEMENTS:

1. Exit the CSATL Lab. [Tag: Room 405]
2. Turn left and walk down the hall, going east.
3. If you are trailing the right wall, you will pass an entrance to a computer lab. [Tag: Room 406]
4. After the alcove, you will detect another hall on your right.
5. Keep going straight.
6. You will pass a glass case on the right wall where they put announcements.
7. After the case, is a large flat-screen tv hanging on the wall.
8. Immediately after the tv, you will find the door to the Computer Science front office on your right. [Tag: Room 414]

The second source route description describes the route from the front office to the restrooms.

ROUTE_ID: R-127
START LANDMARK: Room 414
END LANDMARK: 4th floor restrooms
STATEMENTS:

1. When you exit the office, you are facing north. [Tag: Room 414]
2. Keep walking straight.
3. Immediately across the from the door is the entrance to a small hall.
4. Walk down this hall.
5. When you detect a T intersection, turn left.

6. The men’s bathroom is on the left and the women’s restroom is on the right in a small alcove. [Tag: 4th floor restrooms]

Assuming these two source routes are in the system when a user requests a route description from Room 405 to the 4th floor restrooms, the path inference process will output the following description.

**ROUTE_ID:** Waiting for user approval

**START LANDMARK:** Room 405

**END LANDMARK:** 4th floor restrooms

**STATEMENTS:**

1. Exit the CSATL Lab. [Tag: Room 405]

2. Turn left and walk down the hall, going east.

3. If you are trailing the right wall, you will pass an entrance to a computer lab. [Tag: Room 406]

4. After the alcove, you will detect another hall on your right.

5. Keep going straight.

6. You will pass a glass case on the right wall where they put announcements.

7. After the case, is a large flat-screen tv hanging on the wall.

8. Immediately after the tv, you will find the door to the Computer Science front office on your right. [Tag: Room 414]

9. When you exit the office, you are facing north. [Tag: Room 414]

10. Keep walking straight.

11. Immediately across the from the door is the entrance to a small hall.

12. Walk down this hall.
13. When you detect a T intersection, turn left.

14. The men’s bathroom is on the left and the women’s restroom is on the right in a small alcove. [Tag: 4th floor restrooms]

This route contains an action inconsistency at lines 8 and 9 where the two routes are joined. Because it began R-127, statement 9 describes an initial action of leaving the office. However, in the context of this inferred route description, this action makes little sense since there is no need to actually enter the front office. In this description, the landmark Room 414 is not a destination, but rather is a waypoint that acts as signal for the traveler to make a turn. Therefore, during approval of the description, a user would likely change statement 9 to a completely new statement and delete statement 10 in order to eliminate the action inconsistency.

... 

8. Immediately after the tv, you will find the door to the Computer Science front office on your right. [Tag: Room 414]

9. Turn right, facing north.

10. Immediately across the from the door is the entrance to a small hall.

11. Walk down this hall.

... 

C.3 Example 3

This example shows how heuristic-1 reduces the number of routes used when generating a new route description. All the routes described in this example take place on the first floor of the Taggart Student Center (TSC), a large building on the USU campus, housing many student services and administrative offices. The first source route description in this example describes the route from the Quick Stop, a small convenience store, to the campus post office.
ROUTE_ID: R-461
START LANDMARK: Quick Stop
END LANDMARK: Post Office
STATEMENTS:

1. You are standing with your back to the south entrance to the Quick Stop. [Tag: Quick Stop]
2. Turn left so you are walking east.
3. On your left you will pass the ATM Machines which make distinctive sounds. [Tag: ATM Machines]
4. On the left wall you will find a shelf or counter sticking out from the wall.
5. This is the counter to the Post Office. [Tag: Post Office]

The second source route description describes how to get from the post office to Cafe Ibis, a small coffee shop inside the TSC.

ROUTE_ID: R-141
START LANDMARK: Post Office
END LANDMARK: Cafe Ibis
STATEMENTS:

1. Turn to face east and start walking down the hall.
2. Continue walking east and passing the barber shop, and the copy center as you walk down this long hall.
3. Towards the eastern end of the building, you will come to a wide open area on your left.
4. Turn left and walk a little north.
5. Cafe Ibis is immediately on your left. [Tag: Cafe Ibis]
The third source route describes the route from Cafe Ibis to the patio outside the TSC, which has tables and chairs where students and staff can sit and relax.

**ROUTE_ID:** R-141  
**START LANDMARK:** Cafe Ibis  
**END LANDMARK:** TSC Outdoor Patio  
**STATEMENTS:**

1. After you get your coffee at Cafe Ibis, turn to face the large hall, that is south. [Tag: Cafe Ibis]

2. Walk across until you hit the hall’s south wall which is all windows.

3. Turn left and walk east until you find doors on your right.

4. Exit the door and go through a second door.

5. Walk forward watching our for the big concrete pillars.

6. This is the TSC Outdoor Patio and there tables where you can sit and relax. [Tag: TSC Outdoor Patio]

At this point, a user requests a route description describing a route from the Quick Stop to the TSC patio. If these three source routes are the routes in the system with the lowest cost for this route, the resulting description would be a concatenation of these three routes. The path would consist of a description describing the route from the Quick Stop to the Post Office to Cafe Ibis to the patio.

**ROUTE_ID:** Waiting for user approval  
**START LANDMARK:** Quick Stop  
**END LANDMARK:** TSC Outdoor Patio  
**STATEMENTS:**

1. You are standing with your back to the south entrance to the Quick Stop.[Tag: Quick Stop]

2. Turn left so you are walking east.
3. On your left you will pass the ATM Machines which make distinctive sounds. [Tag: ATM Machines]

4. On the left wall you will find a shelf or counter sticking out from the wall.

5. This is the counter to the Post Office. [Tag: Post Office]

6. Turn to face east and start walking down the hall.

7. Continue walking east and passing the barber shop, and the copy center as you walk down this long hall.

8. Towards the eastern end of the building, you will come to a wide open area on your left.

9. Turn left and walk a little north.

10. Cafe Ibis is immediately on your left. [Tag: Cafe Ibis]

11. After you get your coffee at Cafe Ibis, turn to face the large hall, that is south. [Tag: Cafe Ibis]

12. Walk across until you hit the hall’s south wall which is all windows.

13. Turn left and walk east until you find doors on your right.

14. Exit the door and go through a second door.

15. Walk forward watching our for the big concrete pillars.

16. This is the TSC Outdoor Patio and there tables where you can sit and relax. [Tag: TSC Outdoor Patio]

Since this route description is the result of combining three source route descriptions, there are two action inconsistencies, one at statements 5 and 6, and another at statements 10 and 11. This route could be edited to remove these inconsistencies during user approval. However, suppose the following route is also in the set of source routes at the time the user requests the new route description.
ROUTE_ID: R-147
START LANDMARK: Quick Stop
END LANDMARK: Cafe Ibis
STATEMENTS:

1. Exit the Quick Stop and turn left, walking east. [Tag: Quick Stop]
2. Trail the right wall.
3. You will pass two halls on your right.
4. Next will be doors to the bank, barber, and copy center.
5. After the copy center, there is another hall on your right.
6. Continue east, passing some stairs.
7. When you detect large glass windows on your right, turn left.
8. Walk north, perpendicular to the hall.
9. Cafe Ibis is on the opposite side of the glass windows. [Tag: Cafe Ibis]

With route description R-147 in the set of source routes, there are now two ways to describe a path from the Quick Stop to Cafe Ibis. The first alternative is to concatenate R-461 and R-141. The second alternative is to just use R-147. Because the first alternative consists of two routes, following that path in the digraph will have a higher cost than following just R-147. Therefore, if R-147 is in the system, the new route description would be built from two routes, R-147 and R-141.

ROUTE_ID: Waiting for user approval
START LANDMARK: Quick Stop
END LANDMARK: TSC Outdoor Patio
STATEMENTS:

1. Exit the Quick Stop and turn left, walking east. [Tag: Quick Stop]
2. Trail the right wall.
3. You will pass two halls on your right.

4. Next will be doors to the bank, barber, and copy center.

5. After the copy center, there is another hall on your right.

6. Continue east, passing some stairs.

7. When you detect large glass windows on your right, turn left.

8. Walk north, perpendicular to the hall.

9. Cafe Ibis is on the opposite side of the glass windows. [Tag: Cafe Ibis]

10. After you get your coffee at Cafe Ibis, turn to face the large hall, that is south. [Tag: Cafe Ibis]

11. Walk across until you hit the hall’s south wall which is all windows.

12. Turn left and walk east until you find doors on your right.

13. Exit the door and go through a second door.

14. Walk forward watching our for the big concrete pillars.

15. This is the TSC Outdoor Patio and there tables where you can sit an relax. [Tag: TSC Outdoor Patio]

This route description contains only one action inconsistency occurring at statements 9, 10, and 11 rather than two inconsistencies as were present in the first generated route description. As mentioned during the discussion of heuristic-1, this results in a longer set of statements that was previously acknowledged by users as being safe and understandable. Handling the remaining inconsistency would occur during user approval. The user may decide to retain statement 9 since it offers another waypoint to help guide a traveler. Statements 10 and 11 would be deleted since these actions are no longer required and statement 12 would be edited to remove mention of the turn, also no longer required.

...

9. Cafe Ibis is on the opposite side of the glass windows. [Tag: Cafe Ibis]
10. Walk east until you find doors on your right.

...
CURRICULUM VITAE

John Angus Nicholson, Jr.

EDUCATION

Ph.D., Computer Science. Utah State University, Logan, UT. 2010.


B.S., Computer Science. Purdue University, West Lafayette, IN. 1992.

RESEARCH INTERESTS

Artificial intelligence, assistive technologies, geographic information systems, information extraction, mobile platforms and ubiquitous computing

PUBLICATIONS


INDUSTRY EXPERIENCE


Quality Systems Incorporated. Associate Software Engineer. April 1993 to December 1994


AWARDS


SERVICE


Live demonstration of assistive devices to homeschool students and their parents on the Homeschool Science Day. CS Department, USU. October, 2008.

Live demonstration of assistive devices to homeschool students and their parents on the Homeschool Science Day. CS Department, USU. October, 2007.


Live demonstration of assistive devices to homeschool students and their parents on the Homeschool Science Day. CS Department, USU. February 27, 2006.

Invited as part of the CSATL laboratory by Ann Aust (USU Associate Vice President for Research) to present our lab’s research projects to members of the Utah State Legislature as part of the Legislative Tour of USU. August, 2005.