

# Dialog-Based Location Unaware Wayfinding

*A dissertation submitted*

in Partial Fulfillment of the Requirements

for the Degree of

Btech-Mtech Integrated

by

**Arbaz Khan**

*to the*

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

May 4, 2014

## CERTIFICATE

It is certified that the work contained in the dissertation titled **Dialog-Based Location Unaware Wayfinding**, by **Arbaz Khan**, has been carried out under our supervision and that this work has not been submitted elsewhere for a degree.

---

Prof Harish Karnick  
Computer Science and Engineering  
IIT Kanpur

---

Prof Bharat Lohani  
Department of Geoinformatics  
IIT Kanpur

---

Prof. Stephan Winter  
Department of Geomatics  
University of Melbourne

May 4, 2014

## ABSTRACT

Name of student: **Arbaz Khan**      Roll no: **Y9227128**

Degree for which submitted: **Btech-Mtech Integrated**

Department: **Computer Science and Engineering**

Thesis title: **Dialog-Based Location Unaware Wayfinding**

Names of Thesis Supervisors

1. **Prof Harish Karnick**
2. **Prof Bharat Lohani**
3. **Prof. Stephan Winter**

Month and year of thesis submission: **May 4, 2014**

I plan to finalize the text of the abstract in the end of the writing.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Notions and Terminology . . . . .	5
1.1.1	Wayfinding . . . . .	5
1.1.2	Location Awareness . . . . .	6
1.2	Background . . . . .	7
1.3	Motivation . . . . .	9
1.3.1	Why design a dialog-based system? . . . . .	9
1.3.2	Why choose to be location-unaware? . . . . .	10
1.4	Objectives of the work . . . . .	10
1.5	Roadmap . . . . .	11
<b>2</b>	<b>Problem Definition and Related Work</b>	<b>13</b>
2.1	Dialog Based Wayfinding Systems . . . . .	13
2.2	Evaluation Techniques . . . . .	15
<b>3</b>	<b>Conceptual Model For Wayfinding</b>	<b>17</b>
3.1	Architecture . . . . .	17
3.2	Communication Protocol . . . . .	19
3.2.1	Landmarks . . . . .	21
3.2.2	Simple Intersections . . . . .	21
3.2.3	Complex Intersections . . . . .	22
<b>4</b>	<b>Dialog-Based Localization</b>	<b>25</b>
4.1	Introduction . . . . .	25

4.2	Alternative Reference to Routemarks . . . . .	26
4.3	Extrapolating User Movements . . . . .	27
4.3.1	Estimating User Speed . . . . .	27
4.3.2	Predicting User Speed . . . . .	28
4.4	Detecting Disorientation . . . . .	29
4.5	Reorientation algorithm . . . . .	31
<b>5</b>	<b>Implementation</b>	<b>34</b>
5.1	Dataset . . . . .	34
5.2	Preprocessing . . . . .	35
5.3	Knowledge Base . . . . .	35
5.4	Reorientation . . . . .	35
<b>6</b>	<b>Evaluation</b>	<b>36</b>
6.1	Simulation Setup . . . . .	36
6.2	Goodness Metrics . . . . .	36
6.3	Results . . . . .	36
<b>7</b>	<b>Conclusion and Future Work</b>	<b>37</b>
	<b>Bibliography</b>	<b>38</b>

# Chapter 1

## Introduction

### 1.1 Notions and Terminology

#### 1.1.1 Wayfinding

Human wayfinding is the process of purposeful and directed movement from an origin to a specific destination. It is different from spatial exploration or *locomotion*, the other form of navigation where the goal is not to reach a specific destination, but to contribute to cognitive map formation of an environment. So the daily trips that a person makes to his work-place from his home is a wayfinding task, while exploring an unfamiliar neighborhood in the town is locomotion. The problem of wayfinding is identification of the ordered sequence of actions to be performed in a spatial environment to reach at a desired location. For a car driver in a street network, these set of actions are related to determining the turning behaviour at every intersection that he encounters. For a person in a museum looking for a specific art form, these set of actions would be a sequence of hallways he needs to walk through, to get to the intended location.

Further, its easy to realize that in our everyday interaction with the spatial environment, we are involved with several instances of wayfinding tasks. In some, the exact sequence of actions is well familiar but in others we need either an external assistance or a personal strategy to find our ways. The former case requires acqui-

sition of spatial knowledge of the environment either through prior experiences or through static information learned through maps and other media-based resources and so is prone to errors. Similarly, personal strategies used in wayfinding do not guarantee a successful endeavour to the destination. Traditionally, people have used maps and compasses as external tools as an assistance to wayfinding. These guidance instruments have evolved over the years to mobile navigation systems as the need was to provide incremental delivery of instructions with regards to the movement of the person. This has provably been more effective as the information on what needs to be known is provided only when it needs to be known. This concept shapes the modern form of assistance, *location-aware* wayfinding.

### 1.1.2 Location Awareness

A service is said to be *location aware* if it allows user to discover and communicate their positions in real world using some or the other form of external hardware support. Location awareness has become the key component in any mobile computing application [27]. Talking about navigation, we can say that if the end-device knows its geographical location, the service is location-aware. The definition comes from the realm of location-aware computing where location-awareness means ability to provide services based on the geographical location of a mobile device. Primarily, there are three different techniques of location sensing - Triangulation, Scene Analysis and Proximity Sensing [9]. In the modern day location-aware services, GPS is the most widely used because of its better relative accuracy. The Invisible Ideas Project [24] was the first of its kind to use flash and GPS technology to provide location-aware services.

Having talked about being location aware, its easy to realize the notion of location-aware wayfinding. By the help of techniques of location-awareness, the mobile devices are able to determine the geographical location using a sensing technology (such as GPS) and then using internet, relay the location information (along with any identity-based data like user ID, device ID, etc.) to the service provider.

Thus, it helps in realizing an effective wayfinding assistance system which can provide an incremental delivery of instructions to the user based on the location of the user.

## 1.2 Background

The modern advancements in technology have diminished the efforts involved in discovering and implementing wayfinding strategies. With the advent of smartphones, it has been made possible to build powerful applications which can show maps and compute routes based on preferences. It has removed the complexities of carrying a map and understanding its symbols and notations. The navigation services are dynamic, that is the representation of graphical information keeps changing with regards to the current perception of the user. This is where location-awareness comes into picture and applications have utilized the power of this facility to customize their functionality to maximally benefit the users. The quality of user experience has evolved over a series of developments with offerings like 3-D representation of the environment showing up places nearby for better orientation. Significant efforts have been put into the research and development of the systems for navigational assistance. With the techniques of augmented reality, it has been now made possible to attach digital information (such as images, voice notes) to the environment. Furthermore, local information on weather and traffic have been incorporated into these applications to further facilitate an interactive wayfinding environment.

Despite the powerful services offered as digital navigational aids, there come several issues along with it. When the service is location aware, the end-device needs to have typically high processing speeds to process communicated information. Though the cost and availability of such devices is moving towards the inexpensive side but so are the requirements of these increasingly sophisticated applications. Apart from the installation and usage costs associated with these services, the limited accuracy behind the positioning systems is a vital concern. The inaccuracies extend from measurement errors in positioning systems from satellites to those in the map-



matching algorithms [30] that attempt to associate recorded GPS data points<sup>1</sup> with the correct roadway. The GPS sensors suffer from poor service coverage and need clear vision to the sky to allow locking location. Thus, GPS limitations extend to subways, underpasses, indoor environments and dense street networks with tall buildings. Furthermore, deploying a location sensor like GPS has overheads of cost and power consumption.

The other limitation associated with modern day wayfinding services is their utter dependence on the extrinsic information for providing route assistance. The quality of route instructions are based upon the details in the map and the available landmark information through crowdsourced databases. This availability is fairly non-uniform and the success in providing utility is highly variable. Its easy to find regions where the route instructions in these commercial applications have no incorporation of existing landmarks (due to their unrecorded entry in the spatial databases), and are merely turn-based. Also, despite the availability of the landmark information doesn't guarantee good quality route instructions unless the representational names used are consistent and recognised universally or locally. For example, since not all streets in India have names, various popular mapping services and applications which rely upon street names to convey route instructions, had to invent their own street naming conventions and thus are observably ineffective in the context of navigation assistance.

Beyond these limitations, most of the services rely on the internet connection available between the server and the client particularly for delivering map data or route based information. Streaming such an information requires good internet connectivity. Storing offline maps is an alternative option but it may not be always feasible as and when the space requirements grow beyond accepted limits.

---

<sup>1</sup>These associations are unreliable to an extent and usually mismatch in dense street networks with diverging roadways, overpasses and underpasses

## 1.3 Motivation

While designing a pure auditory-based wayfinding cum location-tracking model would eliminate the high-end device support, here we discuss our underlying motivations behind the two major decisions made before working on the model.

### 1.3.1 Why design a dialog-based system?

In addition to the above shortcomings in modern digital navigation services, research works [13, 22, 25] have identified user preference for auditory assistance as well as a memory advantage for auditory over visual information. It has been observed [1, 6] that auditory guidance facilitates the task of wayfinding whereas visual guidance facilitates cognitive map formation. Ego-centric auditory instructions (i.e., based on the driver's perception) reduce the workload in navigation for drivers who are involved more in the task of route learning rather than cognitive map formation. Besides this, graphical interface has its limitations to applications where interaction with the real world is more important than interactions with the computer (or here, the mobile device). Wayfinding can be treated as one of those tasks that demand high level of attention to avoid any road-side risks. On one hand, these visually dominated applications demand the users attention because they predominantly present information on a digital display (or a mobile screen) and thus are potentially distracting and interfering. While on the other hand, auditory route instructions have been observed to be processed and followed without interference to the driving task [13, 22].

Experiments [29] have also indicated an additional benefit associated with auditory guidance that the reaction times are fastest with pure auditory route instructions as compared to electronic route maps or turn-by-turn displays.

Furthermore, one of the major advantage of a dialog system in wayfinding is that context-identification can be implemented by studying the response of the user to the questions prompted. This helps in personalising the service with regards to individ-

ual navigational strategies preferred by the users. Majority of navigation assistance technologies are still visually dominated and the use of audio is an unexplored area of research. In our knowledge no comparable research has been performed exploiting the usefulness of pure auditory medium for navigation purposes.

### 1.3.2 Why choose to be location-unaware?

On one hand as we speak of benefits behind a dialog-based system for wayfinding devoid of any graphical interface, it may well be said that such a system can be more effectively realized with location positioning systems. Apart from the overheads and limitations of GPS mentioned above, the motivating factor behind the choice of being location unaware is targeting compatibility towards low-end devices. Even though there have been significant advancements in mobile technology and subsequent reductions in the costs of a high-end device, a major portion of third-world countries (such as India) rely upon vanilla feature phones as mobile devices.

Also, we think that a location unaware methodology could be of great use in wireless sensor networks where its infeasible and cost-ineffective to install GPS sensors on every wireless node to determine its location. The idea is that if one can design an effective wayfinding service devoid of GPS support, then it can be applied in developing a messaging protocol for WSNs eliminating the need of location sensors.

## 1.4 Objectives of the work

We target the design of a *Dialog-based Location Unaware Wayfinding* model which exploits the auditory mode of route guidance and eliminates the use of global positioning systems. A service is location-unaware if the end-device has no location sensitivity and can't determine its physical location either by itself (e.g.,GPS) or by sensing the resources from the environment (e.g., Infrared or wireless media). To further define a *location-unaware* service, we describe it as a service under a

pseudo location-aware model which though can determine the location of the end-user within a certain estimate but doesn't employ any hardware support from the end-device for location sensing. The thesis of this work is:

*a dialog-based conversation, inquiring for inferring the physical location of the end-user can achieve pseudo location-awareness.*

Here, we lay down the concrete objectives that we wish to fulfill as a part of this research work.

Firstly, to communicate the route instructions, we aim to design a cross-lingual wayfinding platform which is generic in terms of application to any intended natural language structure. Such a system facilitates customizing the route instructions as per user convenience.

Secondly, we target to build a location-unaware system and attempt to match the effectiveness of positioning systems for localizing a user in wayfinding tasks.

Thirdly, we aim to minimize the dependence of the system on any extrinsic information. By extrinsic information, we mean any information that is not readily and uniformly available due to the needs of large-scale manual processing e.g., crowd-sourcing, or database built from a place directory or gazeteer (gazeteers are not available for all places). The intention is to maximize the use of inherent characteristics of an environment like the height and color of buildings, the existence of an open space and hence its shape, surrounding vegetations, etc. Such characteristics can be automatically extracted from standard detection systems like LIDAR.

Fourthly, we plan to provide a pure auditory-based service which demands nothing more than auditory medium for communication with the mobile end-user. Thus the implication is the service functions independent of support for any wireless technology (such as bluetooth) or web access.

## 1.5 Roadmap

In this thesis, we discuss upon the design of the wayfinding model which uses dialog-based conversations to localize a user and guide him to his destination. To make

the model generic, we design a communication protocol based upon notational representation of route instructions to convey the path to destination. Since there is no device-based support for location awareness, to bring it into effect, an algorithm to generate a structured dialog interaction is architected for tracking the user's location in terms of en-route landmarks encountered. *Chapter 3* introduces the semantics of the communication protocol as well as an architecture of a generic wayfinding model. In *Chapter 4*, we present the localization algorithm to determine user's location by generating prompts for location tracking. It also discusses on the strategy adopted to reorient a user disoriented from his path due to possible misinterpretation of route instructions or because of an erroneous behaviour. *Chapter 5* discusses the implementation of the model which was used to study the effectiveness of the algorithm. To evaluate goodness of the algorithm, a simulation platform was set up to model a user with homogenous and non-homogenous speed patterns and erroneous behaviour in following the route instructions. *Chapter 6* elaborates upon the employed user-modelling and introduces the goodness metrics used for evaluating the results of performance of the route guidance model. We end in *Chapter 7* by providing a brief summary of the contributions made by the work and an outlook on possible improvements of the model.

# Chapter 2

## Problem Definition and Related Work

### 2.1 Dialog Based Wayfinding Systems

Significant research has been done in the area of defining [18] and generating [4, 8, 19, 23] good quality route instructions which are specialized for human understanding. These models have evolved from those working on handcrafted data disseminating turn-by-turn instructions in a one-step-per-sentence mappings to the systems processing data from Geographic Information Systems (GIS) using visible features of the surroundings and generating compact and more human-like descriptions. Yet, most of these systems were monologic systems in which the instructions are provided in advance or even incrementally, causing them to suffer from poor comprehensions and an inability to resolve user's ambiguities along the path. Owing to the benefits of a dialog system to respond to user queries and possibility to react to inadequate or poorly understood instructions, research works [10, 14, 26] had moved towards building dialog-driven wayfinding models. The major concern in these systems is the balance between not overwhelming the user with too much media to minimize the cognitive overload, and to allow several levels of details to produce comprehensive and accurate set of descriptions preventing any ambiguities or confusions. Although

some [26] of these systems rely on purely speech-based interactions while others [10, 14] have used the luxury of realizing a multimodal interface which offers the user a freedom of choice by combining speech, pen and graphics based input signals. Although multimodal systems do help in matching user preferences and result in better and disambiguated input signals, but these systems lose the fundamental requirement of wayfinding assistance, which is, to discourage distractions from driving task and reduce the driver workload. Pure-auditory based route instructions help in servicing systems which cause to minimize the interference to driving task and thus create no negative impact on reaction times [29].

At present, there are only a handful of systems specializing in spoken dialogue-systems for wayfinding. These systems can be classified into three categories depending on the target domain - indoor wayfinding, human-machine interaction and outdoor wayfinding. The systems [5] based upon human-machine interaction aim to facilitate the machine in execution of spatial task by providing route information and handling questions raised by it on detecting ambiguities. These tasks though certainly are similar to real world wayfinding assistance models in dialog structure and end goals but differ in the ability of their subjects. Humans follow subjective and non-uniform speed-patterns, instruction understanding behavior and conceptual memory models.

For outdoor wayfinding, Janarthanam et. al. [12] introduced a mobile dialogue application that combines GIS systems with spoken-dialog interface to assist tourists in city navigation. It allows tourists to initiate tasks such as entity search(e.g. a museum or a restaurant), navigation assistance, entity inquiry and others. The proposed architecture is capable to work in a fairly complex spoken-dialog interaction (e.g., Take me to Hume) by the help of its semantic parser and modules for natural language generation (NLG) and understanding (NLU). Richter et. al. [26] worked on generating context-based route instructions for outdoor wayfinding. Using a spoken dialog interface, the proposed conceptual model allows user to tune the details of the route instructions from incremental turn-by-turn instructions to a more abstract

set of instructions focussing upon destination descriptions. The model identifies and adapts to wayfinder’s spatial knowledge about street network after deciding upon the prominence of a spatial entity derived from a hierarchy of visual, structural and cognitive features. Spoken-dialog systems for indoor wayfinding systems have not been so far realized as per our knowledge. Cuayáhuitl et. al. [3] worked on a text-based dialogue system for indoor wayfinding which can respond to user queries on locations in a building. The authors claimed that a dialogue system for indoor wayfinding with only text-based interaction leads to a very high overall scores of user satisfaction. Despite the fact that indoor settings are perfect for application of location-unawareness, yet none of these systems have yet attempted it and rely upon certain location sensitivity (such as support for infrared and wireless media) of the end-device. Also, the relatively less emphasis on evolution of pure spoken-dialog systems can be understood owing to their limitations in poor speech recognition and unfulfilled linguistic expectations.

In this work we try to overcome these limitations by setting up a cross-lingual communication platform to convey non-ambiguous route instructions. Furthermore, questions posted to the user for location tracking are such that the responses are recognisable speech commands (such as ‘yes’, ‘no’) and require primitive speech processing.

## 2.2 Evaluation Techniques

Simulations in a virtual environment provide an important means for evaluation and improvement of real-world systems. Having set up the virtual environment, the next task is to identify the metrics for evaluating effectiveness of the system. The work done in this thesis is closely related to natural language generation (NLG), only that we focus only on the spatiotemporal content and not the linguistics. Thus, the technique and metrics used for evaluating NLG systems are relevant in this context. For example, if the number of users who reach the target are less in number, it reflects poor quality of NLG involved and thus could be related to the underlying



spatio-temporal content of the instructions. Its only recently that increasing number of efforts have been made for evaluation of NLG systems. As Spanger et. al. [28] point out, there can be two categories of methods to evaluate the effectiveness of a NLG systems - intrinsic and extrinsic. *Intrinsic* methods compare the system with a benchmark/gold standard system, and *extrinsic* methods which use metrics to identify the goodness of the system such as the number of mistakes made by the user while following the instructions [31].

Further, research on evaluation of NLG tasks and investigation of issues moved towards introducing shared task challenges [7]. In this task, the evaluation is assessed with the help of performance of human users who follow the instructions of the NLG in a virtual environment with several rooms and corridors. The performance evaluation uses a set of objective as well as subjective measures. Objective measures include the distance travelled and time taken in reaching the goal as well as the number of instructions provided and the number of words per instruction. While, subjective measures provide a subjective rating of these instructions whether the instructions were useful. An interesting area of research [11] then emerged which works on modelling navigation assistance with real world problems like uncertainty in location and noisy feedback signals from the user. These works exploit the utility of virtual environment to eliminate the time and costs vested in real world experiments as well as provide a means to manipulate the environment targeting specific issues and context.

In this work, we move beyond the above mentioned virtual environments tested on humans. With the virtualization of spatial environment, we also virtualize the user via driver modelling and use a set of goodness metrics to measure the effectiveness of the system. Independent works have been done in the area of simulating driver behavior to study traffic and accident patterns but none of these have been integrated into a testbed for wayfinding assistance.

## Chapter 3

# Conceptual Model For Wayfinding

We introduce our location-unaware wayfinding model through a generic spoken-dialog based model for wayfinding. This model can be transformed to a location-unaware wayfinding model by specializing the module structure. The aim of this model is to assist wayfinding tasks using natural language dialogues. The next section presents an architecture of the model. We then propose a natural language independent communication protocol for conveying route-instructions in wayfinding assistance.

### 3.1 Architecture

The generic wayfinding model is designed to incorporate any linguistic structure and a predefined technique for user localization. It should also be personalized for user and should be able to identify and react to the context realizing the differences and constraints. The knowledge base behind the functioning should be dynamically updateable to incorporate the changes in spatial environment (such as broken road links, new buildings, etc). A pipelined architecture realizing the requirements is shown in Figure 3.1. First, a *user* (real human or modelled) requests for assistance to his destination. The *Natural Language Module* parses the semantics of utterances, processes spatial information and records a new query into the system. Before route instructions could be generated, it is required to locate the user (obtain the source

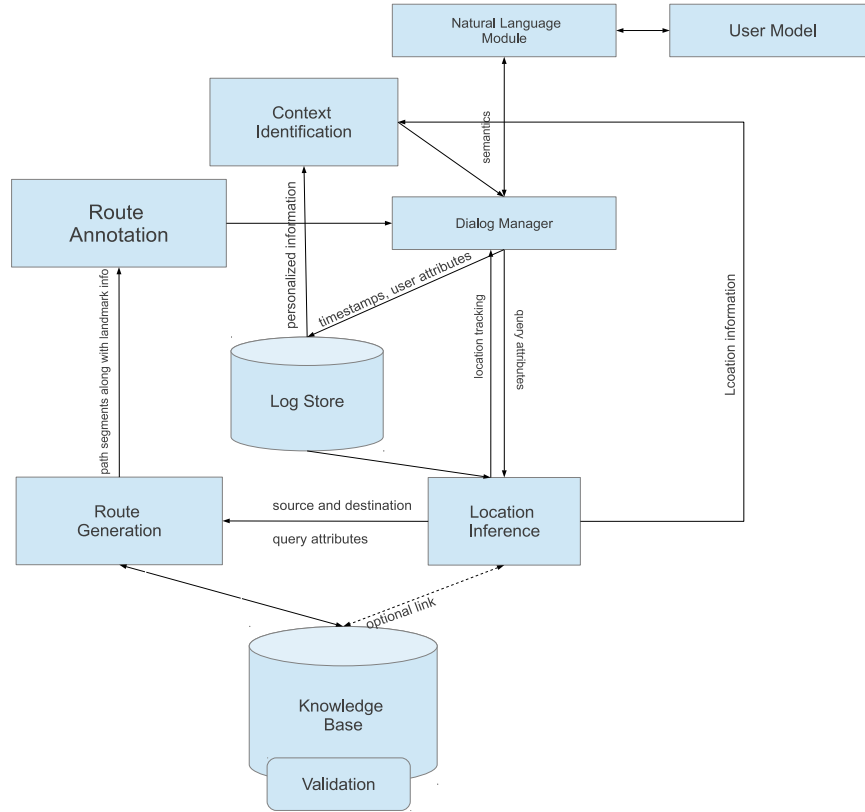


Figure 3.1: Architecture for a generic wayfinding model

location). Most of the wayfinding systems today use GPS systems for this purpose in an outdoor environment and a wireless or infrared media in indoor settings. Baus et. al. [2] introduced an adaptive model for localization that alternates between the different sensing technologies to overcome their individual limitations. For the system introduced in this work, we seek the user's location via dialog-based tracking, posing simple questions using features of a spatial environment. Nonetheless, we can abstract a module specifically for *location inference* which either can work independently (by using sensors or receptors) or uses the *knowledge base* of the system for localization. Thus, the knowledge base stores a characterization of the complete spatial environment. It also stores transportation network for target modalities (e.g., road links for vehicles, abstract spaces for pedestrians) and its relationship with the environment features (e.g., landmark-street association).

Once the information required to compute the topological route is collected, then the desired route is computed using query attributes (such as preferred route option like shortest route, scenic route, etc. and modality constraints like roads permissible

for 4-wheelers). This computed route is passed along with the nearby landmark information for *route annotation*. The route annotator processes the salience of the landmarks, removes redundant spatial relationships and fills up the spatial content of the route instructions. Besides operating on the processed information, the system is also sensitive to the context i.e. user attributes (such as modality, speed patterns) and location information. The context needs to be considered before generating natural language route instructions. With the help of contextual information, a decision is made by the *dialog manager* on the delivery of route instructions to match with the time when a user needs to know it. So, decisions like these make sure that turn instructions are temporally matched with user’s movement. The dialog manager also compiles the route instructions into a form understandable by the natural language module, which then serves to eventually fulfill the user request.

## 3.2 Communication Protocol

In this section we introduce a communication protocol to facilitate in a cross-lingual platform development. The protocol is targeted to deal with two kinds of scenarios - simple intersections and complex intersections. These latter scenario relates with the challenge of dealing intersections with larger number of possible actions (more than 4) and tackling ineffectiveness of standard direction models. but also on In a similar work, Klippel [15] introduced a formal representation of turn directions at decision points which could be chunked to produce better quality route instructions and could be tailored as per user preferences. But the formal theory model targets only simple intersections. Our approach is based upon similar notational representation of turn instructions but extends to complex intersections as well. The goals in designing this protocol are:

1. **Completeness** The protocol should indicate the action to be taken at each decision point implicitly or explicitly. A set of good quality instructions avoids the need of specifying action at each decision point by chunking action be-

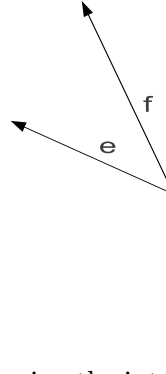


Figure 3.2: Ambiguity arises to determine the intended action when the route instructions are based on standard direction models as *take a left*.

haviour. For example, instructions ‘*go straight*’ and ‘*take a left turn*’ can be compactly presented as ‘*take the second left turn*’ which implicitly directs action behaviour at each decision point and is arguably more comprehensible.

2. **Non-ambiguity** An ambiguous language model leads to confusion in action required at each decision point and thus could lead to disorientation. For instance, instructing to take a left turn in a topology as shown in Figure 3.2 is ambiguous and its not sure whether to take a left at *e* or *f*.
3. **Applicability** Alongwith completeness and non-ambiguity, its equally important to consider the ease of applying the communication protocol in real-world scenario. This means that the protocol should be structured such that it could be translated easily to the desired natural language.

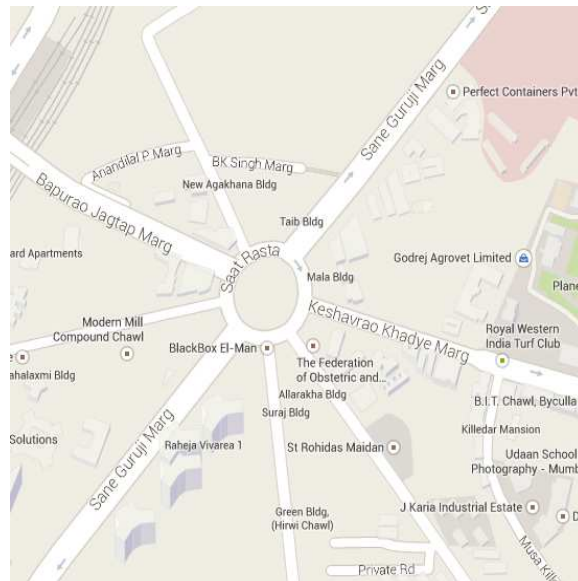


Figure 3.3: Real world example of a complex intersection in Mumbai, India taken from Google Maps [21].

The communication protocol proposed is an adaptive approach to formal representation of route instructions. It uses standard direction model like that of Klippel [15] when the resulting indications are non-ambiguous. However, if the turns are complex, it uses a clock-based convention to represent turn directions. The protocol uses a language independent symbolic encoding for landmarks and turn instructions at decision points. The two scenarios of simple and complex intersections differ in the symbolic encoding of directions while landmarks use a notational representation and each of these is elaborated below.

### 3.2.1 Landmarks

The landmarks are represented by notational IDs and by the direction (left or right) in which user encounters this landmark while moving on a path segment. So, if  $X$  is the notational ID of a landmark and if moving on the directed path segment, the user would encounter  $X$  on his right, then the corresponding representation is  $X^R$ . Similarly,  $X^L$  indicates that moving on the path segment, user can see  $X$  on his left. These representations along with that of turn behaviour at intersections (simple and complex) form the communication protocol.

### 3.2.2 Simple Intersections

Most of the intersections in the real-world fall under this category and are fairly trivial to communicate. To formally define a *simple intersection*, we divide field-view of the navigator in four triangular zones (see Figure 3.4). Choosing reference axis as the left direction perpendicular to the incoming road segment  $p$ , the left zone spans 45 degrees on either side of the reference point. Its mirror image in the field-view w.r.t. the decision point is the right zone. Similarly, one can conceptualize the other two zones. We define a decision point as *simple intersection*, if in each zone of the navigator's field view, there is atmost one road segment emerging from the decision point.

The instructions that are associated with change in direction are represented by

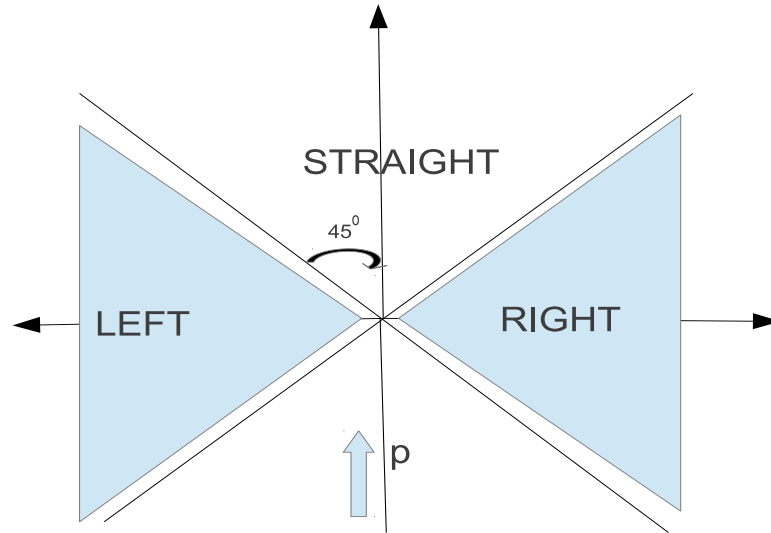


Figure 3.4: A decision point is simple intersection, if in each zone of the navigator's field view, there is at most one road segment emerging from the decision point.

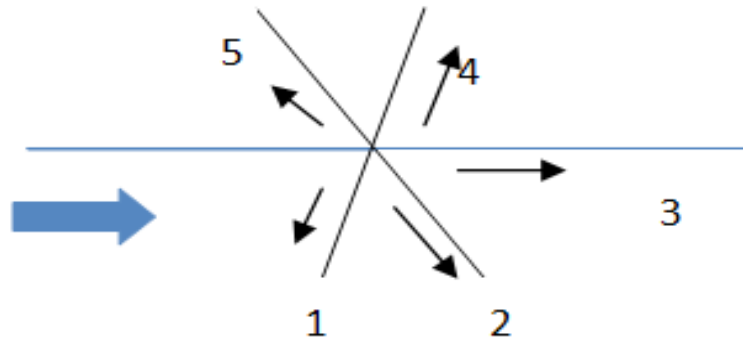


Figure 3.5: Clock based convention to non-ambiguously represent directions at complex intersections

symbols **R** and **L**. The non-turning instruction indicating to go straight is represented by **S**.

### 3.2.3 Complex Intersections

Any decision point which is not a simple intersection is a *complex intersection*<sup>1</sup>. For representing complex intersections, we use a clock-based notation for a non-ambiguous conceptualization. Klippel [15] handles non-standard turns by using 8-sector model which opens up alternate notations such as **hl** (half left), **vr** (veer

<sup>1</sup>See Figure 3.3 and Figure 3.5

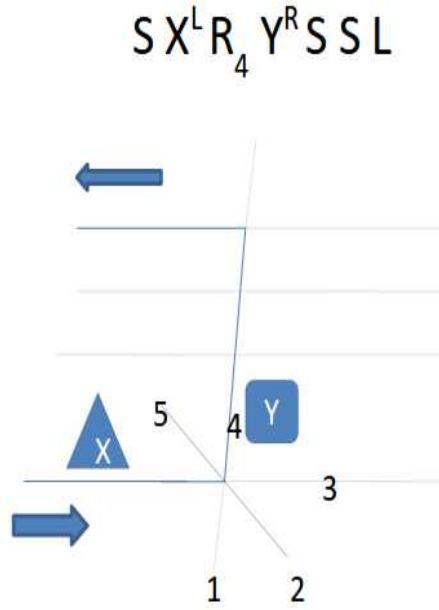


Figure 3.6: An Example to showcase an application of the proposed communication protocol. The intended route is represented symbolically in text at the top of the image. One possible translation of the encoding in english language can be as: 1. *Go Straight* 2. *You would find X on your left* 3. *At the intersection, take the 4th link in anti-clockwise direction from the most adjacent turn at your right side* 4. *You would find Y on your right* 5. *Take the third left after Y*

right), etc. These work well for 4-way intersections and 3-way intersections but as the number of emerging road segments increase beyond 4, the 8-sector model fails as there can be more than one road segment in the same sector.

A clock-based notation doesn't work on the basis of sector model and can handle extended multi-way intersections. The roads are represented in an anti-clockwise numbering starting from the first turn right-adjacent to incoming road segment as shown in Figure 3.5. Notationally, we represent complex turns in the form  $R_i$ , which represents the  $i^{th}$  numbered turn in anti-clockwise direction starting from the most adjacent turn on your right.

Figure 3.6 elaborates a full-fledged example combining the representations of landmarks and intersections and attempts to present an english translation of the same. Before we conclude, we claim that the qualitative calculi used by Klippel [15] on formal representations can be applied likewise to the proposed communication protocol as it is as a direct extension to the theory by including complex intersections. So if a route has no complex intersections, our communication protocol is



semantically similar to that of Klippel. Continuing on the example from Figure 3.6, the route segment corresponding to the representation *SSL* is translated as *the third left* which is similar to the chunking rule of Klippel’s *wayfinding choremes* [15], where by the help of term rewriting, an intermediate representation is deduced from simple representations, prior to natural-language translation.

# Chapter 4

## Dialog-Based Localization

In the previous chapter, we proposed a communication protocol to convey turn behaviour at every intersection. In this chapter, we introduce the algorithm for dialog-based tracking between the intersections to determine user's location and temporally align the delivery of route instructions with the user's movement and confirm his orientation. While posing questions to a user for localization and confirming orientation, we propose the use of an alternative reference to landmarks which highlights the distinctive geometry features. We next present a method to extrapolate movements of a user for location inference using speed predictions. To further cope up with errors and misinterpretations, we also introduce the algorithms to detect and resolve disorientations. The localization algorithm along with the communication protocol are the two mutually exclusive and exhaustive components of the proposed location-unaware dialog-system.

### 4.1 Introduction

While guiding a user to his next path segment, its required to keep track of the user's location to avoid any disorientation. Since, there is no device-based support for location sensing, the only way to localize a user is by asking him about his location in a controlled input format. Though, researchers [16, 17, 20] have been working on processing unrestricted NL input to extract spatial information but

none of these have been able to overcome the classical information extraction errors limiting the output. Further limitations brought by poor speech recognition add to the ineffectiveness of using unrestricted language as input. For this work, we try to limit the questions to those with an objective reply (such as yes or no) or recognisable speech commands (such as color of a building, etc.). These questions are strategically based on the salient features of the spatial environment i.e. *landmarks*. The other requirement to provide quality user interface is to be able to predict the correct position of the user with a good enough estimate, such that the number of questions asked to confirm the orientation and localize the user are kept to minimum. The ‘number of questions asked’, as we would discuss later in detail in Chapter 6, remains as a prime evaluation criteria.

## 4.2 Alternative Reference to Routemarks

In any session of wayfinding, a user is incrementally guided for turn behavior at every intersection. Between every two intersections, user is prompted for localization via questions based on whether he encountered the associated landmarks en-route.<sup>1</sup> Lovelace et al. [18] distinguish between the landmarks according to the purpose they serve i.e. either in choosing the action at a decision point (landmarks) or confirming reorientation on a path segment (*routemarks*). In localization, landmarks are always treated as routemarks and with this, there is a difference in how they are communicated to the user. In the following discussion, the term landmark would be used substitutively for a routemark.

A landmark is referred to by its name (e.g., Eiffel tower) or by its category (e.g., hospital, T-junction) or both, depending on whichever defines its distinctiveness in its locality. Although, in general category-based landmarks are easily comprehensible under route instructions but similar familiarity is not guaranteed with name-based landmarks unless the name is explicitly mentioned in a readable form (like for hotels). In such cases, it is preferable to refer a landmark by its distinctive

---

<sup>1</sup>The issue of how landmarks are associated to a path segment are discussed in Section 5.4.

geometry feature (if any or else choose a landmark by geometry). For example, instead of referring to that building as *Visitor's Hostel*, the system should refer it alternatively as the “red colored 2-floored building” if its the only one so in the locality. Since, even a category-based landmark can be misinterpreted, esp. when its structure doesn't directly reflect its category (like a movie theatre might not look like a theatre), we reference a routemark by its distinctive geometry feature if its salience falls below a certain threshold. The threshold parameter differs for name-based landmarks based on the assumption that the chances of unfamiliarity to a name are more than the mismatch of landmark structure with its category.

## 4.3 Extrapolating User Movements

### 4.3.1 Estimating User Speed

Consider the situation when the user is on a particular path segment and is being guided to his destination through an incremental set of instructions. At this point, the instruction pertaining to the next intersection has been made known. The prompts are such that the user is required to give a positive response only when he sees a particular landmark after taking the needed action. For example, natural language equivalent (in English) to such a prompt could be, *Go straight at the next intersection and prompt me 'Yes' when you see a cafeteria on your left*. These instructions as discussed earlier are a part of the communication protocol. At this point, if the user follows the instruction correctly and takes a non-turning action at the next intersection, he should see the *cafeteria* on his left after some point of time. Since, the instructions do not explicitly or implicitly mention about any intersection in between, it is guaranteed that between the *cafeteria* and current location of the user, there is exactly one intersection. The time when a positive response is recorded from the user, the *cafeteria* becomes the new current location of the user. Thus based on distance measures and the corresponding time difference,

speed can be computed. Thus, speed estimation can be put as:

$$User\ Speed = \frac{Distance\ between\ two\ recorded\ prompts}{Time\ difference\ between\ the\ two\ prompts}$$

### 4.3.2 Predicting User Speed

Speed predictions lead to adaptive nature of the localization algorithm and assist in extrapolating movement patterns of a user as per his speed profile. Continuing from the above example, consider a case when user doesn't respond with a 'Yes' prompt for a relatively long time. If the system acts passively waiting for the 'Yes' prompt, a disoriented user may be totally lost (unless the system accepts unrestricted NL to process a general user query). Thus, there needs to be a predefined time limit in which the user is expected to provide a response. If this time limit expires, the system can interrupt to generate another prompt confirming his orientation. Thus, to get an estimate of this time limit, it is required to predict speed of the user.

The speed predictions are done specific to a road segment and time of day. Some roads facilitate faster speeds than the others but the differences differ over the time of day. Also to be considered is the speed profile of the user. A slow driver might prefer to drive slow proportionately in every speed-limit scenario. Summing up, the algorithm for speed prediction works over two predictors -

- historical average speed on the road segment at this time of day ( $H_{e,t}$ ),
- average relative speed deviation of the user from the historical speeds ( $A_u$ ),
- dynamic slowdown factor ( $f$ ).

For every upcoming road segment, speed is predicted using historical average speeds on the roads and the average deviations of the estimated speed of the user in this session from the historical average speeds on the travelled roads. The feature historical average speeds is stored specific to discretized time slots in a day owing to the observation that a road might be distinctly fast in the late nights as compared to that in peak-hours. The average relative speed deviation feature is related to

the characteristic of the user and can help identify driving preferences. This gives a slow driver enough time gap between successive prompts for confirming orientation, while a fast driver might get different prompts to spatio-temporally match his rapid movements.

The third predictor is meant to consider live traffic information. There may be cases when a road has history of fast vehicle operating speeds and yet, temporary slowdowns maybe observed even for drivers with high-speed driving preferences. Such slowdowns usually occur due to natural interruptions such as foggy weather, torrential rains or snowfall, or even man-made interferences like festival celebrations, rallies or occasional traffic-jams due to miscellaneous reasons. The estimation of dynamic slowdown factor is done freshly at each time slot by averaging relative slowdowns at the given edge from all the users and can be mathematically modelled as:

$$f_e = \frac{1}{|U|} \sum_{u \in U} \left(1 - \frac{(A_u - A_{e,u})}{A_u}\right)$$

$A_{e,u}$  – speed deviation of user  $u$  at  $e$  from historical average speeds at  $e$  at given time slot

$A_u$  – average relative speed deviation of the user  $u$  from the historical average speeds

$U$  – the set of users who travelled the edge  $e$  in this time slot

$f_e$  – slowdown factor for edge  $e$  at given time slot

## 4.4 Detecting Disorientation

As discussed before, exactly one prompt is posed to the user between every two intersection points. This prompt asks the user to confirm his position and is put as early as possible at a road segment to detect disorientation early. There are two cases possible if no response is received from the user within the time limit (set after speed prediction). These cases are handled as below. Algorithm 1 presents the algorithm to detect disorientation.

---

**Algorithm 1:** DetectDisorientation(response,landmark,timeFrame)

---

**input** : Response received at the end of time limit (can be None),  
the landmark position expected and  
current time limit

**output:** 1, if detected disorientation, 0 otherwise

```

if response is positive then
    go to next route segment
    current location becomes this landmark
else
    prompt to ask if the intermediate intersection was crossed
    response  $\leftarrow$  getInput()
    if response is positive then
        timeFrame  $\leftarrow$  timeFrame  $\times$  waitFactor
        // wait for timeFrame
        response  $\leftarrow$  getInput(timeFrame) // blocks for timeFrame units
        if response received and response is positive then
            return 0
        else
            // Here when no response received or negative response
            return 1
    else
        return 0

```

---

## Case-I - Slow Driver

If the user is unable to reach the next position within the expected time frame, then he might not be able to see the instructed landmark before the end of time limit and thus no ‘Yes’ prompt would be received even though the user is not disoriented. Since, it is very much possible to be slow in crossing a intersection esp. if it involves clearing the traffic signals, such cases are duly considered and the system inquires whether the intermediate intersection was crossed. If the answer is ‘No’, then the system delegates the next position to this intermediate intersection and waits for a

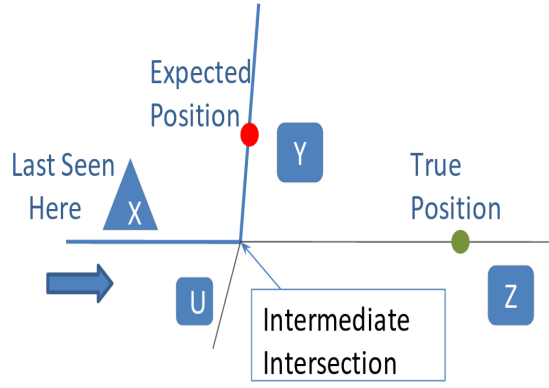


Figure 4.1: **Detecting Disorientation** The highlighted route denotes the path segment communicated to the user for wayfinding. The user has disoriented from the desired path segment and has been detected by the system after a positive response on querying on whether the intermediate intersection was crossed.

‘Yes’ prompt which if received, confirms that the user has crossed the intersection.

## Case-II - Disoriented User

However, if the system receives a ‘Yes’ prompt at the inquiry for crossing the intersection, it waits for an additional time frame (which is sized as that of the original frame reduced by a wait factor). This additional time frame is used to accommodate possible slowness of the driver assuming that the user though running a little slow would encounter the desired landmark at the next path segment. Once the additional time frame expires and yet the user doesn’t see the instructed landmark despite having crossed the intersection, the system traps into the user for reorientation. This situation is depicted in Figure 4.1.

## 4.5 Reorientation algorithm

The process of tackling disorientation is divided into two phases, each with its own characteristic way of causing reorientation.

### Phase-I: Anticipative

The first phase is *anticipative phase* where we estimate user’s position based on movement extrapolation using all possible paths from the last seen point (LSP).



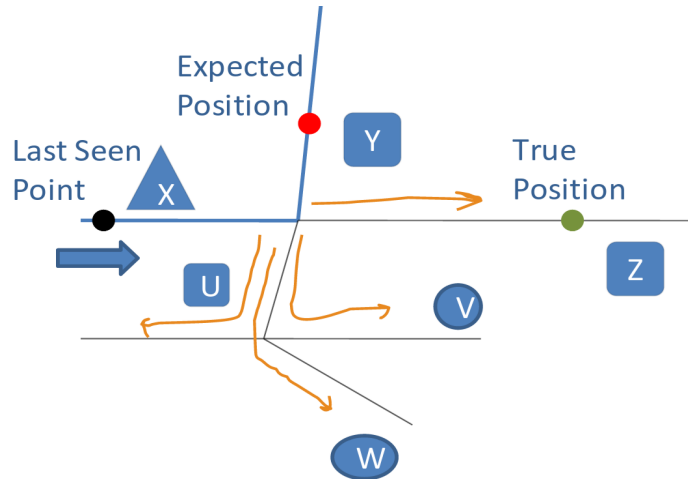


Figure 4.2: **Reorientation:Anticipative-Phase** Prompt the user with questions identifying landmarks  $U$ ,  $V$ ,  $W$  and  $Z$  along with the orientations w.r.t. user (left or right) to get the location estimate. Reorientation is achieved when the user acknowledges any of the identities put in the prompts.

The idea is to localize the user to the nearest landmark visited by the user. In the act of localization, the user is prompted with the possible landmarks he could have encountered en-route. For example, consider the case shown in Figure 4.2. Once disorientation is detected for the user, then based upon speed predictions for the user, the possible path segments and associated landmarks are identified. Thereafter, we prompt the user with questions identifying these landmarks ( $U$ ,  $V$ ,  $W$  and  $Z$ ) along with the orientations w.r.t. user (left or right) to get the location estimate. Reorientation is achieved when the user acknowledges any of the identities put in the prompts.

## Phase-II:Reactive

In a case where small number of prompts suffice in reorientation, one can directly query for identification over all such landmarks seen on the way. But when the number of prompts increases by more than an acceptable extent, we opt to switch the reorientation strategy. Until now, we were being anticipative and the questions asked were of the form ‘*do you see U on your right?*’. In the reactive strategy, we attempt to capture the environment of the user by querying over well-defined attributes. For example, the questions now are of the form ‘*do you see any building*

*nearby?*' The term '*building*' here is one of the categorizing attributes of landmarks which is a part of the feature-set stored as visual features in the knowledge base<sup>2</sup>. The feature-set includes attributes like color of the buildings, heights, shapes of the open spaces and other automatically extractable attributes.

The major drawback in solely relying upon the above strategy is that it doesn't guarantee to localize a user and is dependent upon the stored visual features. In some cases, even after exhaustively asking questions on the categorizing attributes, there still might be a set of possible locations of the user. For example, in an ideally homogenous neighborhood, buildings in all streets might look alike in shape, color and height. For such cases, when there are still a set of locations under consideration, the user is asked to follow his original direction of movement only to be interrupted later for yet another reorientation. This process is iteratively repeated until the responses resolve the location of the user distinctively.

The reactive approach is motivated by the *scene analysis* location sensing techniques [9]. These techniques work on visual images or electromagnetic measurements to sense observed features of a scene for determining a user's physical location. Understandably, the features used are simple that are easy to represent and compare from an observed scene. The approach mentioned is a hybrid of *static* and *differential* scene analysis. In the former technique, features in question are looked up from a pre-defined geo-spatial database, whereas in the latter, differences in scenes observed with user's movements are used to match known spatial environments.

---

<sup>2</sup>See Section 5.4 for more details

# Chapter 5

## Implementation

In chapter 3, we discussed the architecture of a generic wayfinding model, its structure and interaction of the different modules. In this chapter, the implementation aspects of the key components are described with consideration to its application in developing a location-unaware dialog based system. We begin by introducing the dataset for our prototype implementation along with the realized preprocessing procedures. We also discuss the flat-data model used to build the knowledge-base and organized to represent spatial information.

### 5.1 Dataset

The dataset used for implementation of the model was taken from the geospatial data of Indian Institute of Technology, Kanpur. The raw data consists of shapefiles representing different layers of a GIS including road network, geometry and metadata of spatial features such as academic buildings, residential area, parks, etc. Evidently, the IITK map depicts a mix of regions which are variably dense in terms of landmark density. On one hand, the academic area is filled with a plethora of distinct features which can be easily identified by an unfamiliar navigator and residential regions on the other, have almost no salient landmark or mutually distinguishing features. The academic area is built on a dense street network while there are long and clear demarcations in the residential areas.

## **5.2 Preprocessing**

## **5.3 Knowledge Base**

## **5.4 Reorientation**

# Chapter 6

## Evaluation

### 6.1 Simulation Setup

### 6.2 Goodness Metrics

### 6.3 Results

## Chapter 7

### Conclusion and Future Work

# Bibliography

- [1] Carryl L Baldwin. Individual differences in navigational strategy: implications for display design. *Theoretical Issues in Ergonomics Science*, 10(5):443–458, 2009.
- [2] Jörg Baus, Antonio Krüger, and Wolfgang Wahlster. A resource-adaptive mobile navigation system. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 15–22. ACM, 2002.
- [3] Heriberto Cuayáhuitl, Nina Dethlefs, Kai-Florian Richter, Thora Tenbrink, and John Bateman. A dialogue system for indoor wayfinding using text-based natural language. *International Journal of Computational Linguistics and Applications*, ISSN 0976, 962, 2010.
- [4] Robert Dale and Sabine Geldof. Using natural language generation in automatic route description. In *Journal of Research and Practice in Information Technology*, pages 89–105, 2005.
- [5] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3):143–166, 2003.
- [6] Hiroshi Furukawa, Carryl L Baldwin, and Ellen M Carpenter. Supporting drivers cognitive map construction with visual geo-centered and auditory ego-centered guidance: Interference or improved performance. *A. Vincenzi, M. Mouloua, PA Hancock: Human Performance, Situation Awareness and Automation: Current Research and Trends. HPSAA II. D. Mahawa, NJ: Lawrence Erlbaum*, pages 124–129, 2004.

- [7] Andrew Gargett, Konstantina Garoufi, Alexander Koller, and Kristina Striegnitz. The give-2 corpus of giving instructions in virtual environments. In *LREC*, 2010.
- [8] Christopher Habel. Incremental generation of multimodal route instructions. In *Working Papers of the 2003 AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pages 44–51, 2003.
- [9] Jeffrey Hightower and Gaetano Borriello. Location sensing techniques. *IEEE Computer*, 34(8):57–66, 2001.
- [10] Topi Hurtig and Kristiina Jokinen. Modality fusion in a route navigation system. In *Procs of the Workshop on Effective Multimodal Dialogue Interfaces EMMDI-2006. Sydney, Australia*, 2006.
- [11] Srinivasan Janarthanam and Oliver Lemon. The gruve challenge: generating routes under uncertainty in virtual environments. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 208–211. Association for Computational Linguistics, 2011.
- [12] Srinivasan Janarthanam, Oliver Lemon, Phil Bartie, Tiphaine Dalmás, Anna Dickinson, Xingkun Liu, William Mackaness, and Bonnie Webber. Evaluating a city exploration dialogue system combining question-answering and pedestrian navigation. In *Proc. ACL*, 2013.
- [13] Brit Susan Jensen, Mikael B Skov, and Nissanthan Thiruravichandran. Studying driver attention and behaviour for three configurations of gps navigation in real traffic driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1271–1280. ACM, 2010.
- [14] Kristiina Jokinen, Antti Kerminen, Mauri Kaipainen, Tommi Jauhiainen, Graham Wilcock, Markku Turunen, Jaakko Hakulinen, Jukka Kuusisto, and Krista Lagus. Adaptive dialogue systems-interaction with interact. In *Proceedings of*



- the 3rd SIGdial workshop on Discourse and dialogue-Volume 2*, pages 64–73. Association for Computational Linguistics, 2002.
- [15] Alexander Klippel. Wayfinding choremes. In *Spatial Information Theory. Foundations of Geographic Information Science*, pages 301–315. Springer, 2003.
  - [16] T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *5th ACM/IEEE International Conference on Human-Robot Interaction*, pages 259–266, 2010.
  - [17] Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing*, 8(3):4:1–4:36, 2011.
  - [18] Kristin L. Lovelace, Mary Hegarty, and Daniel R. Montello. Elements of good route directions in familiar and unfamiliar environments. In *Proceedings of the International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, COSIT ’99, pages 65–82, London, UK, UK, 1999. Springer-Verlag.
  - [19] Wolfgang Maaß, Jörg Baus, and Joachim Paul. *Visual grounding of route descriptions in dynamic environments*. Univ. des Saarlandes, SFB 314, 1995.
  - [20] Cynthia Matuszek, Dieter Fox, and Karl Koscher. Following directions using statistical machine translation. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, pages 251–258. IEEE Press, 2010.
  - [21] Map of Mumbai. Google maps, <https://maps.google.com/>, May 2014.
  - [22] Dalton P, Agarwal P, Fraenkel N, Baichoo J, and Masry A. Driving with navigational instructions: Investigating user behaviour and performance. *Accid Anal Prev.*, pages 298–303, 2013.

- [23] Thiagarajasarma Pattabhiraman and Nick Cercone. Selection: Salience, relevance and the coupling between domain-level tasks and text planning. In *Proceedings of the Fifth International Workshop on Natural Language Generation, Dawson, PA*, pages 79–86, 1990.
- [24] B. Perry. The Invisible Ideas Project: Using Mobile Devices, Macromedia Flash, and Global Positioning Systems, 2002.
- [25] Ian Reagan and Carryl L Baldwin. Facilitating route memory with auditory route guidance systems. *Journal of Environmental Psychology*, 26(2):146–155, 2006.
- [26] Kai-Florian Richter, Martin Tomko, and Stephan Winter. A dialog-driven process of generating route directions. *Computers, Environment and Urban Systems*, 32(3):233–245, 2008.
- [27] Bill N Schilit, Norman Adams, Rich Gold, Michael M Tso, and Roy Want. The parctab mobile computing system. In *Workstation Operating Systems, 1993. Proceedings., Fourth Workshop on*, pages 34–39. IEEE, 1993.
- [28] Philipp Spanger, Ryu Iida, Takenobu Tokunaga, Asuka Terai, and Naoko Kuriyama. A task-performance evaluation of referring expressions in situated collaborative task dialogues. *Language Resources and Evaluation*, 47(4):1285–1304, 2013.
- [29] Raghavan Srinivasan and Paul P Jovanis. Effect of selected in-vehicle route guidance systems on driver reaction times. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2):200–215, 1997.
- [30] Christopher E White, David Bernstein, and Alain L Kornhauser. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8(1):91–108, 2000.
- [31] R. Michael Young. Using grice’s maxim of quantity to select the content of plan descriptions. *Artificial Intelligence*, 115:215–256, 1999.