Luke Geraghty

Outside the Inside:
Mapping the Steps to a Great Dialogue Agent
For Indoor Route Navigation

Introduction

This paper will critically examine systems that automatically generate indoor route instructions to find properties of a dialogue agent that would best help a user find their way through buildings.

As discussed by Cuayáhuitl et al. (2010), most current route navigation systems are for cars or public transport, with indoor wayfinding being somewhat sidelined. Researcher systems also vary widely in terms of their modality (spoken language, text-based, showing a map), the amount of route navigation steps they give (all at once from a situated immovable agent or incrementally via, for example, a mobile app that allows a user to keep moving), and whether they are restricted to single-floor navigation or can handle stairs.

Five other important issues will also be discussed here:

1) What are some common needs for humans using these systems?
2) Are signs from the environment i.e. landmarks useful?
3) How are the systems evaluated?
4) What are the advantages of an agent that can exchange in dialogue with a human as opposed to a one-way, direction-giving service?
5) Do any systems take into account that some users may have learning/visual difficulties and need a different set of instructions?

A Text-Based Approach

Cuayáhuitl et al. (2010) created a stationary text-based dialogue agent that interacted with students in German and helped them navigate a university building. A user would ask questions via a GUI and be given short simple instructions they could remember or write down if necessary. Here is a translated example dialogue:

```
S: Do you need a route description on this level? Just
   write your question into the text field.
U: Where can I find room number 3180?
S: I have found the following rooms: b3180, a3180. Which
   are you looking for?
U: b3180
S: Please turn around and go straight until the next
   hallway on the left-hand side. Turn left and go until
   the door b3180 on the right-hand side.
   Do you have further questions?
```

This system can be better explained by looking at its pipeline architecture.

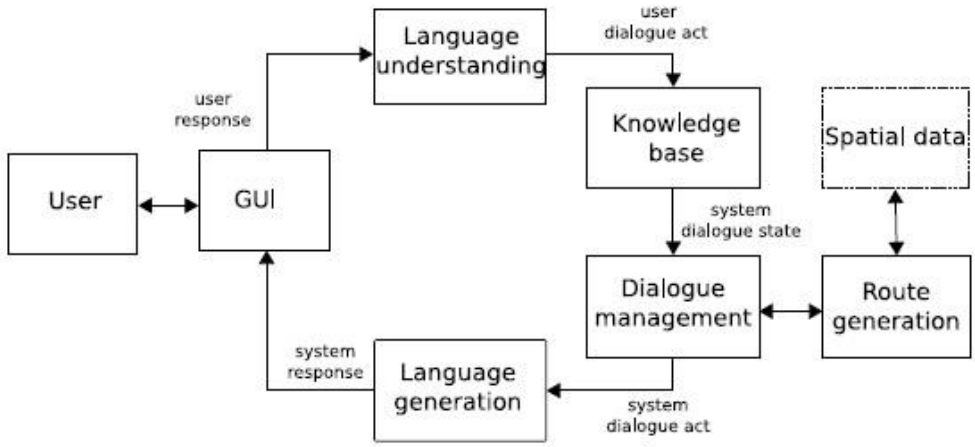

Figure 1: A pipeline architecture for the text-based dialogue system of Cuayáhuitl et al. (2010).

After getting input from a user, a language understanding module parses the text using pre-existing grammars or, if parsing fails, uses keyword spotting to identify i.e. room numbers or the names of people. A dialogue act will be assigned to the input ("ask" for when the user asks a question, for example). Then the dialogue management module decides what the system is going to do by mapping the user's dialogue act to a machine dialogue act held in the knowledge base. These acts could be responses such as requesting more information from the user or clarifying previously given information. The language generation module then provides route instructions by generating logical forms that are given to a language generator, which in turns outputs text to the GUI. Lastly, the knowledge base will update itself as to retain a history of the interaction.

For the generation of route instructions, the system employs a computational process called the Generation of Unambiguous Adapted Route Directions (GUARD) developed by Richter (2008). In GUARD, a route is a graph-like object. Each node on the graph is associated with a decision point: a place where a person will decide what direction they take next (Richter, 2008: 73). GUARD generates context-specific route instructions by taking into account properties of the current environment and also landmarks – a distinct contrast to Internet route planners that disregard both (Richter, 2008: 74). Landmarks will be associated with a decision point based on factors such as distance and potential obstruction, then each landmark tested to see if it can be used as a reference object in the current instruction.

Cuayáhuitl et al. (2010) fed the low level route instructions of GUARD into an algorithm that ultimately generated a more coherent instruction approximating natural human language. Importantly, this algorithm used probabilistic context-free representational underspecification (pCRU) to resolve non-determinacy.

The system was evaluated using twenty-six university students, all native speakers of German. Although a navigator's success is ostensibly easy to evaluate – "did they find the room or not?" – the degree of difficulty a subject had while navigating toward their chosen destination is more difficult to pin down, and there is a lack of standard evaluation metrics for dialogue systems in the wayfinding domain (Dethlefs et al., 2010). Here quantitative measures such as amount of system/user turns and elapsed time were used alongside qualitative measures, for example questions given in a post-performance questionnaire, "Was the system easy to understand?", "Was the pace of interaction appropriate?".

Measure	2 HLIs (52 dialogues)	3 HLIs (52 dialogues)	4 HLIs (52 dialogues)	All (156 dialogues)
Avg. System Turns	2.25	2.38	2.28	2.30
Avg. User Turns	1.30	1.61	1.64	1.52
Avg. System Words per Turn	34.05	40.04	49.59	41.30
Avg. User Words per Turn	4.06	5.34	4.84	4.79
Avg. Time (in seconds)	20.69	19.77	25.87	22.14
Parsed Sentences (%)	23.8	4.3	22.5	16.7
Spotted Keywords (%)	74.6	91.4	73.2	79.9
Unparsed Sentences (%)	1.6	4.3	4.2	3.4
Binary Task Success (%)	96.2	100.0	88.5	94.9
3-Valued Task Success (%)	92.3	88.5	63.5	81.4
4-Valued Task Success (%)	94.9	92.3	75.6	87.6
Easy to Understand	4.65	4.6	4.08	4.46
System Understood	4.71	4.62	4.62	4.65
Task Easy	4.60	4.54	3.73	4.29
Interaction Pace	4.71	4.65	4.52	4.63
What to Say	4.71	4.63	4.65	4.66
System Response	4.60	4.62	4.58	4.56
Expected Behaviour	4.64	4.50	4.21	4.45
Future Use	4.46	4.37	4.12	4.31
User Satisfaction (sum)	37.1	36.5	34.5	36.0
User Satisfaction (%)	92.7	91.2	86.3	90.0

Table 1: Average evaluation results for Cuayáhuitl et al.'s (2010) wayfinding system. HLI stands for High-Level Instruction, a human-like sentence generated by the system and given to the user for navigation. For example, "Turn left and go straight until door B3180 that is at your right." One complete route may be made up of multiple HLIs.

Table 1 shows that interactions between users and the computer were short, that parsing was generally unsuccessful and spotting keywords a vital failsafe, and that the longer the instructions, the lower the overall success and user satisfaction.

In evaluation of this system, it can be said that clearly a text-based dialogue system can work for giving out indoor route navigation. There are often many different (and not always grammatical) ways to ask for a room, hence grammars may struggle. This system cannot move with the user either. They have to remember the steps, which may be long, or take notes.

The authors also say that user-machine interactions are short, as highlighted by Table 1, but that is because this system is limited to one floor only – it cannot do multi-floor navigation of a building, which is an extremely common need of users of such systems. When you add in navigation via lifts, stairs, or both, route navigation is likely to change a great deal, and perhaps a purely text-based approach without a generated map will be less than adequate.

Furthermore, this text-based system takes no user accessibility needs into account. The "one size fits all" approach is not reflective of modern day life. Users of such a navigation system may be cognitively challenged, visually impaired, disparaging of technology, or have any number of needs that the young well-abled university students used in the evaluation might not have had. All test subjects were also native speakers of the system's language, although it should be noted that the system is language independent (Cuayáhuitl et al., 2010: 299) and thus its language accessibility could be improved.

Lastly, perhaps on an obvious note, this system uses no speech whatsoever. The user has to manually type in their answer. If one thinks of a typical university building user, it is quite possible to imagine a student with a coffee cup in one hand and a phone in the other. A multimodal system might be more suited to the needs of the average 21st century university user.

Navigating on the Move

Clearly a navigation system that is rooted to the spot is going to be limited in its ability to give continual help to a user. One quarter of the globe are now smartphone users (eMarketer, 2014). A dialogue agent that is contactable through a downloadable app would allow users to stay mobile while giving updated advice based on their current location.

Unfortunately, very few such dialogue agents exist. Yet it can still be useful to look at a more typical one-way, direction-giving service that is not a dialogue agent to highlight the advantages and disadvantages of such services.

Russo (2013) developed an Android-based application called IndoorNav that helped people find their way in a building of the Delft University of Technology in the Netherlands. One of the key problems in developing a system that wants to keep track of its user is localisation or finding where its user currently is (Russo, 2013:19). GPS cannot be used for indoor navigation because radio signals cannot pass through solid walls. The author also did not wish to use indoor radio signals from, for example, WiFi networks due to all-too-common signal impairment and the desire to have a system that could work in emergency cases without power so long as the user had a network connection with data (Russo, 2013: 20).

IndoorNav instead works with QR codes. A QR code is a barcode that can be read by a machine and encoded with more information than a typical UPC barcode. When booting up IndoorNav, a user is asked to scan a nearby QR code to get their current location, and can update where s/he is by scanning a subsequent code.

Figure 2: A QR code is scanned for use by IndoorNav, allowing it to find the user's current location. Such codes were distributed widely through the university building used in Russo's (2013) study.

Advantages of a system using QR codes are low operating costs and exact positioning not always afforded by WiFi, which has difficulty assessing a user's current floor and suffers from *multipath propagation*: the phenomenon whereby a device that receives radio waves will receive signals from a large number of different paths, often because of reflective interference, which causes this received data to be malformed or lost entirely (McClaning, 2012: 189). QR codes do unfortunately cause the user to bear the burden of updating his/her own location, and downgrade Russo's (2013) system to using a static positioning system. It is this author's view that they may very well grind a navigation in progress to a halt by forcing a user to change from searching for a room to searching for a QR code. This is not necessarily a trivial task when you are unfamiliar to a building, or a new user of an app, or in a rush to get to a university room before the start of your next lecture.

Once a user has scanned a code, they enter the desired end location and whether they wish to use elevators or not. This request will then be sent to a web service which sends an XML response that is parsed and rendered in the next view as a set of textual instructions.

Use of IndoorNav is best demonstrated in a step-by-step fashion (see next page).

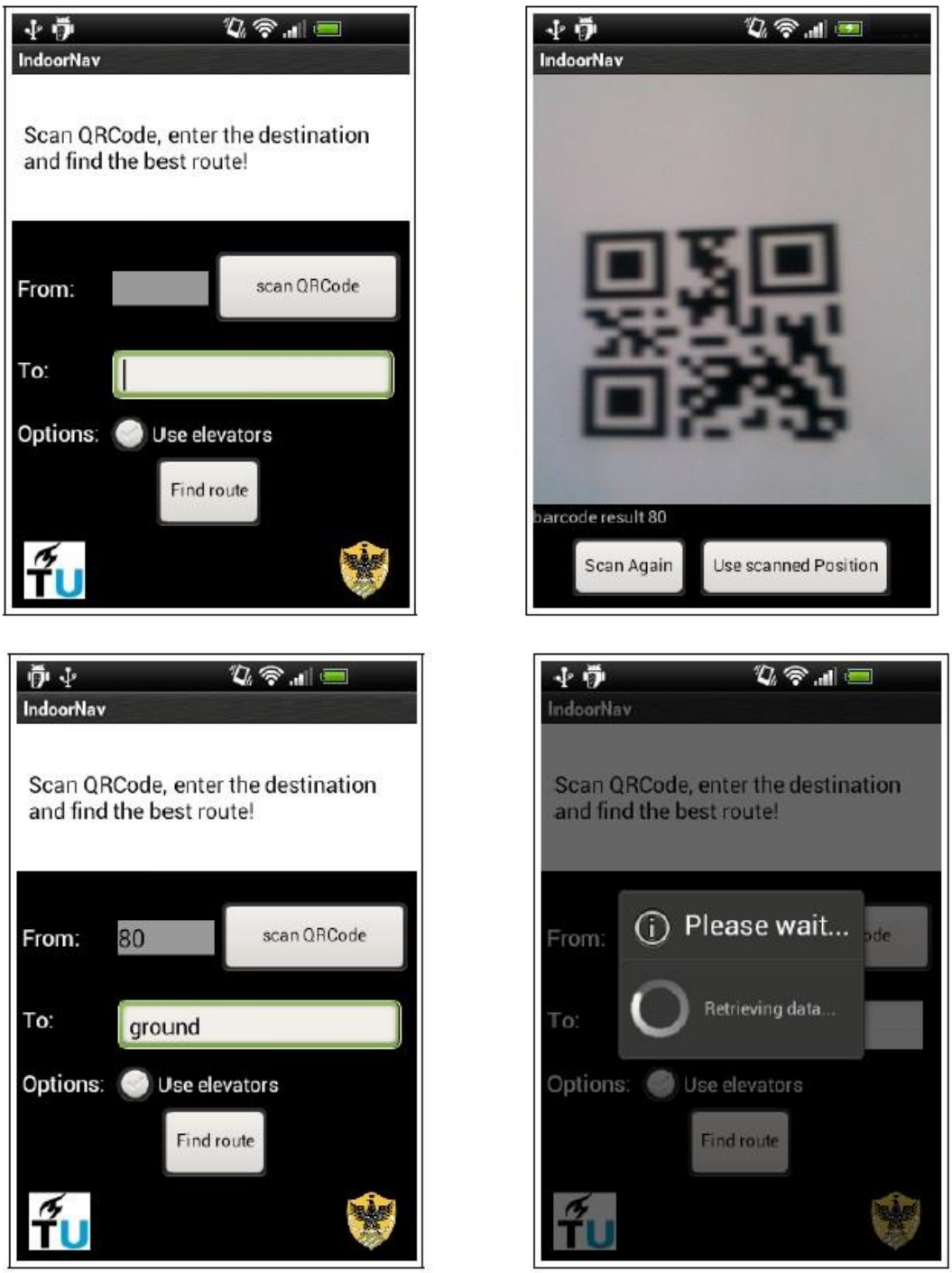

Figure 3: (Clockwise from top left)
1) A user starts the IndoorNav app.
2) They scan a QR code.
3) They enter in their target location.
4) And they send off their request to a web service.

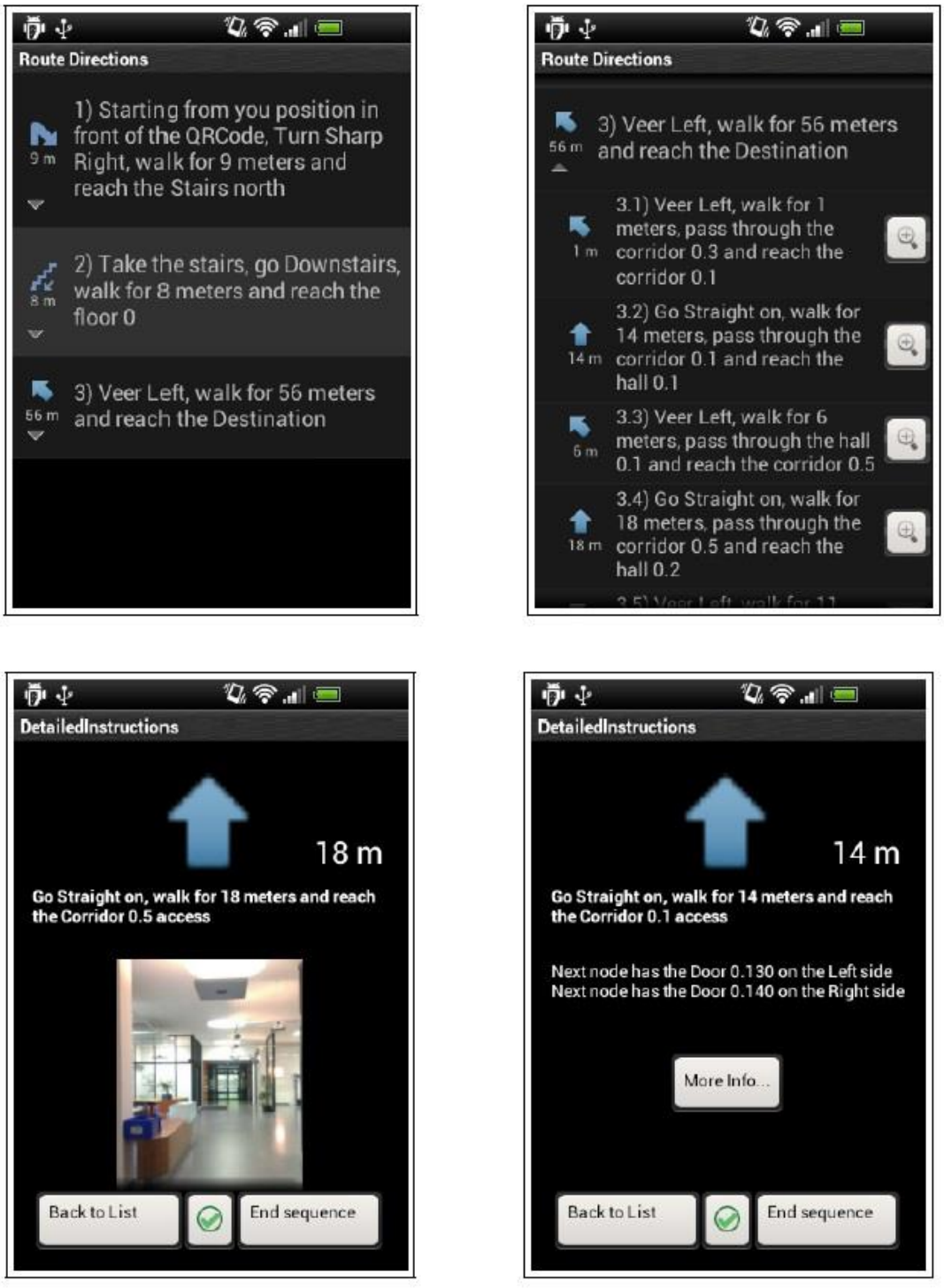

Figure 4: (Clockwise from top left)
1) A user then receives route directions. Note: a change in floor can be handled by this system.
2) Each step of these directions can be clicked upon to see more fine-grained steps…
3) And each of these steps optionally comes with a image which is associated with the next target node…
4) Or additional information such as adjacent door numbers.

Unfortunately, Russo's (2013) own evaluation was limited, giving five students five different route tests for a total number of twenty-five tests. The success rate was 100% and included floor changes. Russo found that floor directions – general directions that ignore room changes and exact amount of distance travelled – were generally sufficient for a user to navigate to their destination.

That said, unlike Cuayáhuitl et al. (2010), the evaluation metrics here were reasonably poor, with no quantitative data such as how long users took to navigate and no qualitative data like that obtained from the questionnaire the aforementioned study gave to participants.

This system also differs by reducing the route navigation problem from a conversation to a foregone conclusion. There is no ability to field questions from the user as with a dialogue agent. Rather, it is assumed the system has provided everything they will need in one go.

On the flip side, this system is another example of one that uses a building's landmarks i.e. an unusual painting to help users find their way. IndoorNav uses GUARD to detect potential landmarks and then picks those that are route-specific. Landmarks have been found to be paramount to successful human navigation (Marchette et al., 2015, Sorrows and Hirtle, 1999: 37).

Clearly prototypical indoor navigation services think in terms of spatially large goal destinations like rooms, not giving users the opportunity to find specific objects in a building. They also use descriptions that do not make use of the same metrics as a human-produced utterance would. How useful or natural is the description of distance in an utterance such as, "Go Straight on, walk for 18 meters"?

The Benefits of a Robotic Situated Agent

A robotic agent will have to refer to objects and places in its environment so as to guide humans. Yet there are many possible ways a robot may refer things in its world.

Consider the following examples of possible referring expressions from Zender and Kruijff (2007):

1) "the location at position (X =5.56,Y =−3.92,q=0.45)"
2) "the mug to left of the plate to the right of the mug (...)"
3) "Peter's office no. 200 at the end of the corridor on the third floor of the Acme Corp. building 3 in the Acme Corp. building complex, 47 Evergreen Terrace, Calisota, Planet Earth, (...)"
4) "the area"

Although valid descriptions, each referring expression fails to achieve its communicative goal, providing too much information, too little, or providing it in such a way that does not approximate real human speech on the same subject.

As Zender and Kruijff (2007: 2) point out, a conversational robot, if it hopes to surpass the indoor navigation service previously described, should use qualitative descriptions in a similar manner to humans, generating referring expressions that distinguish spatial entities in large-scale space. An additional benefit of robotic systems like that developed by Zender and Kruijff (2007) is that they allow humans to search buildings for objects, not only people or rooms. The ontology-based conceptual map within the robot's spatial representation meant that relations existed between areas and objects.

Another advantage of a robot dialogue agent over a static route-giving service is that the robot can be taught new information about indoor spaces. Shi and Krieg-Brückner (2008: 33) highlight how it is always possible that a robot's spatial representation may be out of sync as spaces are constantly changing, hence why it is necessary to allow new route instructions to goals which weren't previously known to a robotic agent.

Lastly, there are algorithms for use in robotic agents that take note of personal and stylistic differences. Oßwald et al. (2014) modelled the problem of giving route directions as a reinforcement learning problem in terms of a Markov decision process, developing an algorithm that learned how to provide good route descriptions from a corpus of human-written directions. One of the features of the directions given to the algorithm was called *description style*: information given by participants that differed person to person (whether they used cardinal directions, allocentric directions, directions relative to landmarks, street names, mileages etc.). Perhaps it is possible for a robotic agent generating directions to be similarly sensitive to style.

Putting It All Together: An Ideal Dialogue Agent for Indoor Route Navigation

An ideal dialogue agent would:

- Use human-like referring expressions.
- Adjust based on user.
- Use text, voice, and maps.
- Use landmarks native to the area, and filter these landmarks according to the route under discussion.
- Be multilingual.
- Operate across floors.
- Be able to continue giving advice (in the case of static i.e. as-you-enter museum information screens) via an app that could locate and guide the user.

Conclusion

Right now the field of indoor navigation systems is highly interdisciplinary, with engineers working alongside geographers and psychologists, all trying to understand how indoor space differs from outside space and the challenges humans face when navigating the former compared to the latter. There is, however, a need for linguists to be involved in this discussion. Even non-robotic services that use dialogue could benefit from an understanding of what it is to ground or clarify – complex behaviours that cannot be easily hand sewn into systems (Cuayáhuitl et al., 2010: 300).

Obstacles still remain for developing a great dialogue agent for indoor route navigation. Buildings are not roads. Neither is navigation by walking the same as that by car. Traditional services on the horizon, such as Google's Indoor Maps, face the same problem as dialogue agents: a lack of easily accessible building information, with Indoor Maps being reliant on user-uploaded floor plans. With many buildings being privately-owned spaces, it seems unlikely that a single company will be able to map the indoor world.

Bibliography

Cuayáhuitl, H., Dethlefs, N., Richter, K. F., Tenbrink, T., and Bateman, J. 2010. A Dialogue System for Indoor Wayfinding Using Text-Based Natural Language. *International Journal of Computational Linguistics and Applications, Volume I.* pp 285-304

Dethlefs, N., Cuayáhuitl, H., Richter, K. F., Andonova, E., Bateman, J. 2010. Evaluating Task Success in a Dialogue System for Indoor Navigation. *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue* pp.143-146

eMarketer, 2014. 2 Billion Consumers Worldwide to Get Smart(phones) by 2016. Available from: http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694. [Retrieved: 11th January 2016].

Marchette, S. A., Vass, L. K., Ryan, J., Epstein, R. A. 2015. Outside Looking In: Landmark Generalization in the Human Navigational System. *The Journal of Neuroscience*, *35(44)* pp. 14896-14908

McClaning, K. 2012. *Wireless Receiver Design for Digital Communication*. 2nd edition. SciTech Publishing pg. 189

Oßwald, S., Kretzschmar, H., Burgard, W., and Stachniss, C. 2014. Learning to Give Route Directions from Human Demonstrations. *Proceedings of the IEEE International Conference on Robotics & Automation (ICRA).* Hong Kong, China. pp. 3303-3308

Richter, K. F. 2008. *Context-Specific Route Directions – Generation of Cognitively Motivated Wayfinding Instructions*. IOS Press, Amsterdam, The Netherlands.

Russo, Davide. 2013. Route Directions using Visible Landmarks for an Indoor Navigation System based on Android device: “IndoorNav". Master’s thesis, University of L'Aquila.

Shi, H. and Krieg-Brückner, B. 2008. Modelling Route Instructions for Robust Human-Robot Interaction on Navigation Tasks. *International Journal of Software Informatics, Vol.2, No.1*. pp. 33-60

Sorrows, M. E. and Hirtle, S. C. 1999. The Nature of Landmarks for Real and Electronic Spaces. In Freksa, C. and Mark, D. M., eds. *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science (COSIT ’99)*. Stade, Germany: Springer. pp. 37-50

Zender, H. and Kruijff, G. J. M. 2007. Towards Generating Referring Expressions in a Mobile Robot Scenario. *Language and Robots: Proceedings from the Symposium (LangRo'2007)*. Aveiro, Portugal. pp. 101–106.