5-2022

Feasibility studies in indoor localization through intelligent conversation

SHESHADRI SMITHA
Singapore Management University, smithas.2022@phdcs.smu.edu.sg

Linus CHENG

Kotaro HARA
Singapore Management University, kotarohara@smu.edu.sg

Feasibility Studies in Indoor Localization through Intelligent Conversation

Smitha Sheshadri
smitha3shesh@gmail.com
Singapore Management University
Singapore, Singapore

Linus Cheng
linus_cheng_@hotmail.com
Singapore University of Technology and Design
Singapore, Singapore

Kotaro Hara
kotarohara@smu.edu.sg
Singapore Management University
Singapore, Singapore

Abstract

We propose a model to achieve human localization in indoor environments through intelligent conversation between users and an agent. We investigated the feasibility of conversational localization by conducting two studies. First, we conducted a Wizard-of-Oz study with $N = 7$ participants and studied the feasibility of localizing users through conversation. We identified challenges posed by users' language and behavior. Second, we collected $N = 800$ user descriptions of virtual indoor locations from $N = 80$ Amazon Mechanical Turk participants to analyze user language. We explored the effects of conversational agent behavior and observed that people describe indoor locations differently based on how the agent presents itself. We devise "Entity Suitability Scale," a concrete and scalable approach to obtain information to support localization from the myriad of indoor entities users mention in their descriptions. Through this study, we lay foundation to our proposed paradigm of conversational localization.

CCS Concepts

• **Human-centered computing → Natural language interfaces**; *Empirical studies in HCI*.

Keywords

conversational agents, indoor human localization,

ACM Reference Format:
Smitha Sheshadri, Linus Cheng, and Kotaro Hara. 2022. Feasibility Studies in Indoor Localization through Intelligent Conversation. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491101.3519617

1 Introduction

Every day, we use location-based services (*e.g.,* Google Maps) to know our positions with respect to our surroundings and to search for routes to reach our destinations. However, these services become unusable in indoor environments due to their dependency on localization technologies like GPS. The lack of sufficient awareness about where we are within complex indoor environments is not just an inconvenience, but could even cause anxiety [17], especially because people lose orientation more easily indoors [7]. Previously proposed sensor-based approaches can localize people in indoor environments precisely, but their dependency on the sensing infrastructure, the high cost of deployment, incompatibility with existing devices, and latency in localization impede their wide adoption [11].

In this paper, we investigate the feasibility of *conversational localization.* We ask, *"Can we identify a user's position within a complex indoor environment by asking them to describe their surroundings as if to a friend or colleague?"* We aim to obviate the need for sensing infrastructure by delegating effort in sensing the surrounding environment to the user; the agent could then focus on processing the user-provided information to compute their position. Prior research has studied the use of verbal instructions to navigate people [6, 7]. But instead of navigation, our work focuses on indoor localization—the necessary but often-overlooked precursor to implementing wayfinding services.

We conducted two preliminary studies to investigate the feasibility of conversational localization. First, we conducted a Wizard-of-Oz study with N=7 participants to investigate how people interact with the hypothetical conversational agent when involved in wayfinding and localization tasks within a university campus. The study result showed that the agent (*i.e.,* an experimenter disguised as a chatbot) could deduce users' positions with user-provided information containing location cues. But we also observed that users interact using description-rich and unstructured language and mention a wide variety of indoor entities. This suggested that the agent must process such raw information and dynamically decide what indoor entities are suitable for localization.

Thus, to further analyze the language people use and to categorize the various indoor entities people mention, we conducted a study on Amazon Mechanical Turks (AMT). We collected natural language descriptions of different indoor locations using 360° tours of multiple locations within a university campus from N=80 participants. We employed a factorial design to investigate the impact of modifying certain agent behaviors on user descriptions. This study uncovered interesting findings about interaction between agent features that presented several implications for the design of the conversational agent. The findings suggest that people can be guided to identify more useful information. We developed and used a Named-Entity-Recognizer (NER) to extract the entities in user-provided descriptions. We then further classified them into different tiers of indoor entities based on their suitability for localization

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519617

and devised "Entity Suitability Scale"—a scoring system to dynamically categorize and group indoor entities based on their structural, visual, semantic, and functional suitability for localization.

2 Related Work

Conversational Agents. Conversational agents, more popularly known as chatbots, have grown prevalent [2]. Utilizing the recent advent of technologies to interpret and generate natural language, researchers and designers have created an array of conversational applications, such as improving education and learning experiences [9, 12, 29], providing social and emotional support [13, 19], and customer service [32]. These applications set forth a promising premise for the effectiveness of chatbots. We thus see an opportunity to extend what today's location-based services offer to users through conversational technology. However, the research domain remains under-explored despite the theoretically well-supported potential in using natural language in human wayfinding [24, 27].

Sensor-based indoor localization. To locate a target's position in the indoor environments, numerous studies have proposed solutions that utilize sensors combined with wireless signals (*e.g.,* Wi-Fi [4, 22, 30, 34], Ultra-wide Band (UWB) [21, 23, 33], FM signal [31], Infrared [16, 35], Ultrasound [1], and RFID tags[18]). These solutions use strategically-placed signal access points across the area of interest. Through the system installation process, one needs to fingerprint the strength of the signal in each unit of the partitioned environment. The system then estimates the location of a sensor-carrying target using the signal strength received by the sensor or measuring signal transmission and reception timings. While these approaches achieve high localization accuracy, implementation overheads associated with hardware deployment and maintenance and labor-intensive on-site surveying impede their adoption.

Instead of utilizing actively emitted signals and sensors, prior work has also worked with images collected from camera phones [14, 25]. For instance, Ravi *et al.* [25] made users wear their smartphones around their necks as pendants and collected images at certain time intervals. Their system compared the collected images to stored images with known locations for localization. However, this solution still suffers from the overhead of collecting and geographically tagging images. More recently, Li *et al.* [14] designed a system that localizes users by asking them to capture short videos. The system processes video clips and automatically identifies visual signage, and estimates the locations of the corresponding points-of-interest. While using visual landmarks to locate the user position is related to our approach, we hand off the responsibility of identifying landmarks to humans. People are better at interpreting visual scenes than an artificial system; as our research shows below, humans can not only provide signage but are also capable of providing other useful landmarks (*e.g.,* presence of a room).

Routing Instruction Generation. Prior work in architectural and urban design sought to understand the roles and significance of visual landmarks in navigating people [6, 7, 20]. Duckham *et al.* [6] studied ways in providing landmark-based route instructions to support people's wayfinding in unfamiliar outdoor environments, and Fellner *et al.* [7] adapted the approach for indoor environments and highlighted the importance of distinctive "landmarks" in wayfinding. Ohm *et al.* [20] studied visual saliency of various indoor objects using eye-tracking to identify the objects that grab human attention during route-based indoor navigation. However, the prior work considered only the situation where the user and the system already know the user's position. Unlike the prior work, we focus on localization, which is a precursor to other location-based applications like navigation. Our work learns and incorporates the significance of visual landmarks in conversational localization from this body of prior work.

3 A Feasibility Evaluation of Conversational Localization

Wizard-of-Oz Study Method. We conducted a Wizard-of-Oz study to assess the feasibility of localizing a person in the indoor environment through a conversation. We recruited (N=7) local university students unfamiliar with the indoor study site. Upon arrival at the study site, we gave the participants three tasks.

i. *Navigation Task*: We instructed each participant to navigate from a starting point to a destination. The experimenter physically guided a participant to the starting point and left them alone. The participant then interacted with the conversational agent (*i.e.,* the experimenter) via chat on Discord [5]. The task was complete once the participant reached the destination by requesting and following navigational instructions from the agent.

ii. *Localization Task*: After the Navigational Task, the participant was asked to head to a random location within the same building. Upon reaching the location, the participant initiated a conversation, and the experimenter attempted to locate their position. The task simulated the situation where the user does not know their location within a building. We instructed the participant not to explicitly state the room names/numbers to make the task realistically hard. The task was complete when the agent identified the location based on their interaction.

iii. *Localization + Navigation Task*: The participant was then instructed to move to another random location of their choice in the same building. The agent identified the participant's location through the conversation, just like the Localization Task, then supplied a set of navigational instructions to guide the participant from their current location to the study site, like the Navigation Task.

The experimenter followed a script (Appendix A) as much as possible to imitate the hypothetical chatbot and resorted to using their judgement only in case the participant behaved unexpectedly. The experimenter used indoor entities in participants' descriptions and floor maps to estimate the participant's location.

Result. All the participants were able to complete the three tasks, and the experimenter successfully located the participants in Localization tasks. We collected 410 chat utterances between the participant and the experimenter disguised as a chatbot. $N = 215$ were participants' utterances. In completing each task, participants made 8.3 utterances on average ($min = 5$, $max = 14$) for the Navigational Task, 7.7 utterances ($min = 3$, $max = 14$) for the Localization Task, and 14.9 utterances ($min = 7$, $max = 27$) for the combined task.

From the 215 participant utterances, we manually identified 106 unique indoor entities. Participants used various phrases to describe

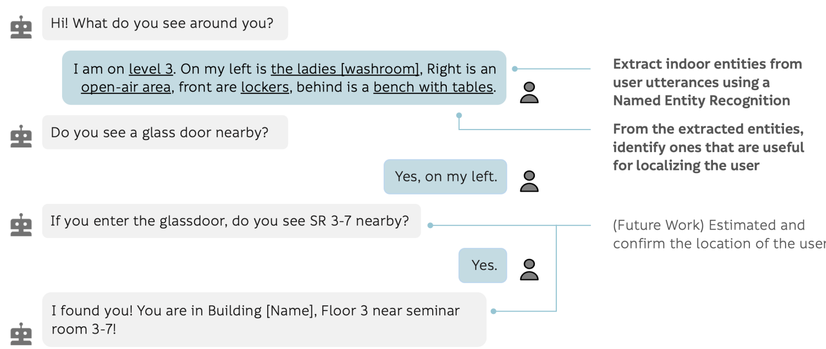

Figure 1: An excerpt of the interaction between an agent (*i.e.,* experimenter) and a participant from the Wizard-of-Oz study (Section 3). The human experimenter could locate a user using an indoor map and entities extracted from a conversation. We envision a conversational agent system that automatically extract indoor entities from user utterances and locate their indoor position.

these indoor entities (emphasized in the following examples). For example, to refer to an open stretch hallway in the study area, participants used "open area" (P2) and "passage that leads to the toilet and other classrooms" (P3). We observed that participants struggled to describe when there was a lack of distinct or familiar entities in their surroundings. They would resort to mentioning whatever they saw and could name, *e.g.* "In front of me is a wall" (P1) and "to my right is a deadend" (P7). The experimenter thus needed to use their judgement and knowledge about the study site in addition to floor maps to disambiguate the entities.

To this end, we developed a custom named-entity-recognizer (NER) in Python using the NLTK library [15] to process the user descriptions by analysing the participant descriptions from Wizard-of-Oz study. The NER removed all punctuation from the descriptions and converted them to lower case. It then tokenized the descriptions using NLTK's `regex_tokenizer` and tagged the tokens with part-of-speech (POS) using `NLTK.tag.pos_tag`. For example, the sentence "I am outside near some empty benches" got converted into `[(i, NN), (am, VBP), (outside, IN), (near, IN), (some, DT), (empty, JJ), (benches, NNS)]`, where the second element of each tuple represents a POS tag (*e.g.,* NN: a singular noun). The script then compared the token sequence to the grammar that we defined; we devised the grammar rules iteratively by trial-and-error. Using the grammar, we created a `RegexParser` instance to parse and chunk the identified tokens into terms or the "named entities". See Appendix B for more implementation details. We describe the evaluation of the NER in section 5.1 below.

3.1 Measuring Entity's Suitability for Localization

Indoor entities described in a conversation vary in their usefulness for localization. Consider the following description, *"To my right is a room [...] with glass doors. Behind me is a pillar in the middle of the hallway [...]"* (from a Section 4 study participant). The NER extracts entities `["room"]`, `["glass" ,"doors"]`, `["pillar"]` and `["hallway"]`. As humans, we instinctually perceive a variability in the "usefulness" of the different entities. *E.g.,* the entity "room" would arguably be more useful in localization compared to "hallway". That is, we make a judgement about which pieces of information to use for localization and what to ask next in order to refine our estimation of the target's location. This indicates that (i) entities could be categorized into tiers based on their suitability for indoor localization, and (ii) an autonomous system should be able to rank the entities by their suitability so that it can utilize the best entities to localize people.

We compiled the following four dimensions that characterize this "usefulness" of entities—which we call "Entity Suitability Scale"—by adapting existing literature in architecture, urban planning [17], indoor [7] and outdoor localization [6]:

i. *Signage-based imageability*: The visual characteristic of a useful entity should separate itself from its environment and let the user identify it with ease. Such a characteristic is referred as "imageability" [17] or "prominence" [7]; highly imageable entities possess distinguishable visual, cognitive, semantic, and structural elements [28]. Examples of prominent entities include signages like room number plates and restaurant signs.
ii. *Permanence*: A structure's permanence affect its suitability as a localizational cue. An entity which occupies the same space within a floor map for a reasonably long period of time can be considered as a permanent. Permanent entities contribute to localizational knowledge whereas temporary entities without a dedicated location do not [6].
iii. *Spatial Extent*: Narrower the area that the entity occupies, more precise the information it provides for locating a target. For example, a "shop" can contribute a more precise location than an entire "wall" [6, 7].
iv. *Ease of Mapping*: For an agent to utilize entities to localize a user, it has to know the entities' locations. That is, we need to be able to identify entities in data sources like floor maps and feed them into the system's locational database. Thus, we consider whether the entities can be found on floor maps as one of the dimensions.

In the study described below, we use this scale to investigate what entities people mention and how useful they are for indoor localization. Note, although some features may be more critical in assessing entities' suitability, we treat the importance of each

Feature	Question	If yes,	Else,
Signage-based Imageability	Does the entity have signage?	1	0
Permanence	Is the entity built into the infrastructure or is it connected to something which is built into the infrastructure?	1	0
Spatial extent	Is it a decision point?*	1	0
Ease of mapping	Is it on the available floor map?	1	0

Table 1: Entity Suitability Scale used to assess usefulness of indoor entities for localization. We take a sum to compute entity's suitability score. (* A decision point is entity where a navigational decision can be made. *e.g.* Turn right at the ATM.)

feature equally. We come up with a if/then rule for each feature, and assign 1-point to an entity if it satisfies the condition (see Table 1). We then sum up the points to calculate the total suitability score of the indoor entity in question. A higher score on Entity Suitability Scale thereby indicates better suitability for indoor localization.

4 AMT Study Method

We conducted an online study on Amazon Mechanical Turk (AMT) to: (i) evaluate our NER script's accuracy in extracting location entities, (ii) investigate how people use language to describe unfamiliar indoor environments, and (iii) assess suitability of the entities that people mention.

We recruited N=80 participants from AMT to describe indoor locations presented in 360° panoramic images. We turned to remote crowdsourcing study due to the current COVID-19 restriction which prevented us to conduct an in-situ study. For the study, we captured 166 360° images from 24 locations across the university campus with an average of six images ($min = 5$, $max = 8$) per location. The participants used a web browser interface to virtually look around the indoor environment of ten locations via pan-and-zoom interaction—see Appendix C for the task interface. We also investigated the differences in descriptions based on type of indoor location and on how the agent presented itself. We adjusted the instructions and types of indoor images to study whether how we ask the participants to provide the information affected their responses. More specifically, we considered the following three manipulation variables:

i. Presence of *Spatial Signage* (*with-signage* vs *without-signage*). Presence of salient visual landmarks could affect how people describe the indoor environment. We exposed half of our participants to indoor images with unique signage (*e.g.,* restaurant sign) (*with-signage* group) and the other half to images without such signage (*without-signage* group). We selected 10 images with unique signage and 10 images without signage for the tasks.

ii. Disclosure of the *Agent Context* (*friend* vs *bot*). Prior chatbot research reported people offered more information when they thought they were talking to a fellow human than a computer agent [26]. To study if this verbosity holds in our setting, we instructed a half of our participants to imagine that they are describing the indoor locations to a *friend* and the other half to think they are describing for a *bot.*

iii. Guidance using *Example* Descriptions (*with-example* vs *without-example*) We studied the effect of providing an example indoor environment description on the elicitation of entity information. For half of our participants, we provided an example description (*with-example*) and no such example for the other group (*without-example*). In the instruction for the *with-example* group, the following "good" and "bad" responses were added: *"I am near Pastamania. It's next to a food court and I can see a shop called Dim Sum. I can also see an escalator."* (good) and *"Very like."*, *"I felt very happy and very involved with this image."*, and *"indoor shop"* (bad).

To avoid confound due to order effect [8], we opted for a 2×2×2 between-subjects design study where we recruited 10 participants for each condition. We recruited N=80 participants who were adult U.S. residents and had an assignment acceptance rate ≥ 99% on AMT with no additional eligibility criterion. We reimbursed US$5 upon task completion.

5 AMT Study Result

5.1 Participant description processing with NER

Our floor maps contained 16 labels corresponding to entities in the location of the tours used for the AMT study. One research team member manually searched and identified 566 occurrences of the 16 entities in the 800 descriptions. We also ran our NER on the descriptions and extracted 99.11% (=561/566) of the manually identified entities (*i.e.,* recall = 0.9911). The NER missed five of the manually tagged entities, four of which contained typographical errors and one contained unusual phrasing (*"a restaurant that serves rice"* instead of naming the restaurant). Note that, we did not use an off-the-shelf entity tagger [10] as it missed many entities with proper nouns in their names ($recall$ = 46.28% or 262/566). In addition to entities on the floor maps, the NER also extracted various other indoor entities mentioned by participants. Overall, the NER extracted 4368 phrases from the 800 descriptions. The shortest entity was "TV" and the longest was "long white wall many different brown doors". We removed all duplicates and retained 2195 unique phrases. People over-mentioned and our NER over-detected entities, necessitating us to rank them by their suitability. See Section 5.3 for the analysis.

5.2 Analysis of language and entities in participants' descriptions

The participants recognized and described 1,229 entities (that are automatically extracted) in areas *with-signage*; this is 56% (=1229/2195) of all the entities, suggesting that participants mentioned fewer entities in locations *without-signage*. The difference in the number of entities mentioned looked more prominent in the absence

Figure 2: (a) A dot plot of each condition’s entity count. (b) Interaction between *spatial signage* and *agent context* factors. People tend to mention more entities in the *friend* context when signage is present. (c) Interaction between *spatial signage* and *example* factors. There is a trend where people mention less entities when signage is absent and the instruction does not provide examples.

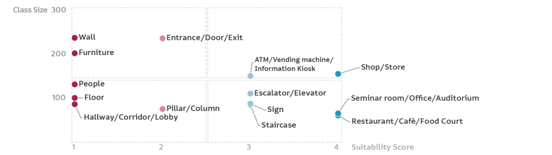

Figure 3: A scatter plot showing entity suitability score-entity class sizes relationship—raw data in Appendix B. Entities in classes like “Shop/Store” are oft-mentioned and suitable for localization. Classes like “People” are rarely mentioned and not useful.

of example. The participants in the *with-example* group identified slightly more entities (52.3% = 1149/2195) than *without-example* group. The participants in the *bot* group mentioned slightly fewer entities (48.4% = 1062/2195) compared to the *friend* group.

We used Generalized Linear Model (GLM) with a log link function to study whether there are significant differences in entity counts between conditions. As our object of analysis was frequency data, we refrained from using methods like ANOVA that assume continuous data that follows a normal distribution. The result indicated significant main effects of agent context (*i.e., bot* vs *friend*) ($\chi^2(5) = 46.23, p < 0.001$) and significant interaction between the spatial signage and agent context at $p < 0.05$. We also observed a possible trend in interaction between the spatial signage and example descriptions at $p < 0.1$. Due to the presence of interactions, we refrained from discussing the main effects. The significant interaction between spatial signage and agent context (Fig 2.b) indicated that participants mentioned more entities under friend context in areas with signage. This partly supported previous findings [26] that people shared more information with fellow humans. But the increase seemed to be conditioned on the presence of prominent signs in the environment. The possible interaction between spatial signage and examples (Fig 2.c) suggested that providing examples are helpful in increasing the number of entities in the responses particularly in areas without signage.

5.3 Suitability of Entities for localization

Participants mentioned a variety of indoor entities ranging from precise locational entities such as shops and rooms with names to ambiguous entities such as “glass doors” or “potted plants”. Quantifying the suitability scores of all 2,195 entities is a gruelling task. We thus turned to a three-step semi-automated process to calculate the entities’ suitability: (i) we first converted the entity phrases to comparable numerical vectors using embeddings; (ii) we then clustered the entities into indoor entity classes; finally, (iii) we manually calculated suitability score for each indoor entity class.

i. Creating Embeddings. The entities that we extracted using the NER script are a set of terms; to use an unsupervised clustering algorithm to group the entities in the next step, we convert each term into a vector of real values (*i.e.,* an embedding). We created embeddings of the entities using the Universal Sentence Encoder [3], which is suitable for encoding sequence of words like the extracted entities.

ii. Unsupervised Clustering. We clustered the entities represented as embeddings using k-medoid algorithm; through trial-and-error, we decided to use $k = 18$ for the number of clusters. We then studied the 18 clusters and removed four clusters that consisted of typos or were solely populated with descriptive parts-of-speech such as adjectives, retaining 14 clusters with 1806 unique entity phrases in total—*i.e.*, the indoor entity classes.

iii. Manual Suitability Score Calculation. For each indoor entity class, one member of the research team manually calculated the suitability score, following the process described in Section 3.1. (See Appendix D) As a result, each class had a score between 1 and 4, where 1 is less suitable and 4 is more suitable for informing the user’s indoor location.

To understand the frequency of entities that people mention and how useful that information is in localization, we draw a scatter plot where each dot represents an indoor entity class (Fig 3); the x-axis represents the entity class’s suitability score and the y-axis represents how often the participants mentioned entities in each class (See Appendix E for size of each entity class). We split the chart into four quadrants to make the following observations: (i) entity classes like “Shop/Store” and “ATM/Vending Machine” are often-mentioned and useful (top-right quadrant); (ii) entity classes like “Room” and “Restaurants” are rarely mentioned but would be useful for localization if mentioned (bottom-right), (iii) people often mentioned things like “Wall” and “Furniture”, but they are less useful (top-left); and (iv) some entities are rarely mentioned and not useful (bottom-left).

6 Discussion and Conclusion

We demonstrated the feasibility of localizing a person in an indoor environment in a conversational manner through our Wizard-of-Oz study. We identified challenges in processing natural language to extract information useful for localization. We developed an NER with 99% recall and devised the Entity Suitability Scale to measure suitabilty for localization. Our work sets forth a foundation for the design of the conversational localization system.

Design Implications. Our AMT study result indicates the potential to improve elicitation of entity information from users by manipulating agent behavior. The absence of signage saw less number of entities in user descriptions. Our system needs to find ways to nudge people provide information that is useful in areas with less signage; guiding people on how to describe the indoor environment using examples could be effective. Our result also suggested the importance of anthropomorphism to the chatbot design; to elicit useful information for localization, it would be important to communicate with the user so that they would feel they are talking to a "friend" rather than an inorganic autonomous agent. The future conversational system could use the suitability score combined with the frequency of the entity class to inform subsequent lines of conversation. For example, if people mention entities which have lower suitability scores but are mentioned frequently (*e.g.,* Wall), the agent can warn users about the lower suitability using tutorials or tips. On the other hand, we also take note of highly suitable entities which are rarely mentioned (*e.g.,* Restaurant) and could take steps to guide users towards them. We anticipate this approach to be useful in indoor environments that possess entities that are easy to spot and familiar for us to name (*e.g.,* universities with uniquely numbered rooms, shopping malls with legible signage, airports with gates with unique numbers) [17].

Limitations and Future Work. The Entity Suitability Scale, while in congruence with our intuitive understanding of "usefulness" of entities, needs to be vetted with more rigorous testing in the future. More analysis is also needed in estimating the spatial boundaries of entities. Our work focused on investigating the feasibility of localizing a person in the indoor environment using the indoor description provided by a user, but we did not study if an automated chatbot can handle a natural conversation with a user to organically elicit information that is suitable for localization yet. In the future, we intend to develop and study the localization capacity of the conversational localization system on-site.

Acknowledgments

This research is supported by Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant.

References

[1] Pascal Bihler, Paul Imhoff, and Armin B Cremers. 2011. SmartGuide–A smartphone museum guide with ultrasound control. *Procedia Computer Science* 5 (2011), 586–592.

[2] Petter Bae Brandtzaeg and Asbjørn Følstad. 2017. Why people use chatbots. In *International conference on internet science*. Springer, 377–392.

[3] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Céspedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175* (2018).

[4] Zhenghua Chen, Han Zou, Hao Jiang, Qingchang Zhu, Yeng Chai Soh, and Lihua Xie. 2015. Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization. *Sensors* 15, 1 (2015), 715–732.

[5] Accessed December. 2021. Discord Instant Messaging Platform. https://discord.com/

[6] Matt Duckham, Stephan Winter, and Michelle Robinson. 2010. Including landmarks in routing instructions. *Journal of Location Based Services* 4, 1 (2010), 28–52.

[7] Irene Fellner, Haosheng Huang, and Georg Gartner. 2017. "Turn left after the WC, and use the lift to go to the 2nd floor"—Generation of landmark-based route instructions for indoor navigation. *ISPRS International Journal of Geo-Information* 6, 6 (2017), 183.

[8] Darren Gergle and Desney S Tan. 2014. Experimental research in HCI. In *Ways of Knowing in HCI*. Springer, 191–227.

[9] Bob Heller, Mike Proctor, Dean Mah, Lisa Jewell, and Bill Cheung. 2005. Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 3913–3918.

[10] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). To appear.

[11] Dheryta Jaisinghani, Rajesh Krishna Balan, Vinayak Naik, Archan Misra, and Youngki Lee. 2018. Experiences & challenges with server-side wifi indoor localization using existing infrastructure. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 226–235.

[12] Alice Kerlyl, Phil Hall, and Susan Bull. 2006. Bringing chatbots into education: Towards natural language negotiation of open learner models. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, 179–192.

[13] Dongkeon Lee, Kyo-Joong Oh, and Ho-Jin Choi. 2017. The chatbot feels you-a counseling service using emotional response generation. In *2017 IEEE international conference on big data and smart computing (BigComp)*. IEEE, 437–440.

[14] Mingkuan Li, Ning Liu, Qun Niu, Chang Liu, S-H Gary Chan, and Chengying Gao. 2018. SweepLoc: Automatic video-based indoor localization by camera sweeping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–25.

[15] Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028* (2002).

[16] Ren C Luo, Ogst Chen, and Pei Hsien Lin. 2012. Indoor robot/human localization using dynamic triangulation and wireless Pyroelectric Infrared sensory fusion approaches. In *2012 IEEE International Conference on Robotics and Automation*. IEEE, 1359–1364.

[17] Kevin Lynch. 1960. *The image of the city*. Vol. 11. MIT press.

[18] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. 2003. LANDMARC: Indoor location sensing using active RFID. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE, 407–415.

[19] Kyo-Joong Oh, Dongkun Lee, Byungsoo Ko, and Ho-Jin Choi. 2017. A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 371–375.

[20] Christina Ohm, Manuel Müller, Bernd Ludwig, and Stefan Bienk. 2014. Where is the landmark? Eye tracking studies in large-scale indoor environments. (2014).

[21] Timothy Otim, Alfonso Bahillo, Luis Enrique Díez, Peio Lopez-Iturri, and Francisco Falcone. 2020. Towards Sub-Meter Level UWB Indoor Localization Using Body Wearable Sensors. *IEEE Access* 8 (2020), 178886–178899.

[22] Jun-geun Park, Ben Charrow, Dorothy Curtis, Jonathan Battat, Einat Minkov, Jamey Hicks, Seth Teller, and Jonathan Ledlie. 2010. Growing an organic indoor location system. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*. 271–284.

[23] Alwin Poulose and Dong Seog Han. 2020. UWB indoor localization using deep learning LSTM networks. *Applied Sciences* 10, 18 (2020), 6290.

[24] Jennie E Pyers, Anna Shusterman, Ann Senghas, Elizabeth S Spelke, and Karen Emmorey. 2010. Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences* 107, 27 (2010), 12116–12120.

[25] Nishkam Ravi, Pravin Shankar, Andrew Frankel, Ahmed Elgammal, and Liviu Iftode. 2005. Indoor localization using camera phones. In *Seventh IEEE Workshop on Mobile Computing Systems & Applications (WMCSA'06 Supplement)*. IEEE, 1–7.

[26] Nicole Shechtman and Leonard M Horowitz. 2003. Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 281–288.

[27] Anna Shusterman, Sang Ah Lee, and Elizabeth S Spelke. 2011. Cognitive effects of language on human navigation. *Cognition* 120, 2 (2011), 186–201.

[28] Molly E Sorrows and Stephen C Hirtle. 1999. The nature of landmarks for real and electronic spaces. In *International conference on spatial information theory*. Springer, 37–50.

[29] Rainer Winkler and Matthias Söllner. 2018. Unleashing the potential of chatbots in education: A state-of-the-art analysis. (2018).

[30] Chenshu Wu, Zheng Yang, Yunhao Liu, and Wei Xi. 2012. WILL: Wireless indoor localization without site survey. *IEEE Transactions on Parallel and Distributed Systems* 24, 4 (2012), 839–848.

[31] Jiang Xiao, Kaishun Wu, Youwen Yi, and Lionel M Ni. 2012. FIFS: Fine-grained indoor fingerprinting system. In *2012 21st international conference on computer communications and networks (ICCCN)*. IEEE, 1–7.

[32] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3506–3510.

[33] Yuan Xu, Yuriy S Shmaliy, Yueyang Li, and Xiyuan Chen. 2017. UWB-based indoor human localization with time-delayed data using EFIR filtering. *IEEE Access* 5 (2017), 16676–16683.

[34] Chouchang Yang and Huai-Rong Shao. 2015. WiFi-based indoor positioning. *IEEE Communications Magazine* 53, 3 (2015), 150–157.

[35] Dan Yang, Weihua Sheng, and Ruili Zeng. 2015. Indoor human localization using PIR sensors and accessibility map. In *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE, 577–581.