Workshop Proceedings of the 9th International Conference on Intelligent Environments
J.A. Botía and D. Charitos (Eds.)

doi:10.3233/978-1-61499-286-8-635

Smart City Search: A User Survey

Richard MIETZ [a,1], Kay RÖMER [a]
[a] *Institute of Computer Engineering, University of Lübeck, 23552 Lübeck, Germany*

Abstract. The number of sensors deployed in cities is growing from year to year due to different drivers such as sensors contained in embedded systems, smartphones, weather stations, and sensor networks. When connecting these sensors to the Internet, the state of the city is available online and in real-time, thus, realizing the vision of an Internet of Things (IoT). The combination of this sensor knowledge with data from other sources and services makes cities "smarter" and enables novel applications such as finding free parking spots. A key service for many of those applications is a search engine. This *Smart City Search* would allow users to search and find entities of the real-world which exhibit a certain state at query time. To inform the design of such a search engine, we conducted a user survey to identify usage patterns and desires of users for *Smart City Search*.

Keywords. Internet of Things, Smart City, Search, Sensor, Survey, Questionnaire

Introduction

Through technological progress the number of sensors in our environment is constantly growing. Modern smartphones contain, among others, sensors for acceleration, light, sound, and position. Both, the commercial and crowdfunding communities started to develop and sell additional sensor equipment for mobile phones to measure environmental conditions such as temperature, humidity, or different gas concentrations. Additionally, several projects and companies are working on sensors for the household to, e.g., detect presence or accidents of people. Finally, researchers deploy sensor networks in cities. One of the most noteworthy efforts is the SmartSantander project [2] where 20.000 sensors will be deployed and connected to the Internet over four cities of which 12.000 are set up in the Spanish city of Santander. These trends will facilitate an Internet of Things (IoT) where entities of the real world are connected to the Internet and publish their state in real-time. In [1], ICT infrastructure such as the IoT is, among others, characterized as a pillar for so-called Smart Cities. "Smart" means that combining data from sensors and other sources will allow a multitude of new applications and services which can "[...] fuel sustainable economic growth and a high quality of life, with a wise management of natural resources [...]" [1]. For example, SoundOfTheCity [2] is a smartphone app which measures the sound intensity around the user with the microphone and uploads it to a central server. The collected noise levels of all users are displayed on a map. In this way, users can find, e.g., quiet places to relax and authorities can, e.g., use the data for urban development to reduce noise. Some other imaginable scenarios are pictured in the

[1] Mail of Corresponding Author: mietz@iti.uni-luebeck.de
[2] http://www.smartsantander.eu/

IoT comic book [3] which was supported by the IoT-i Project. As this example illustrates, searching entities with a certain state will be an important service for the Smart City. So, instead of searching only for the static homepage of a restaurant, one could search for an Italian, well-tempered restaurant which has free tables at the moment. The infrastructure for such *Smart City Search* must be, on the one hand efficient in terms of resource consumption because this is the limiting factor of most sensor systems such as smartphones and sensor nodes, but, on the other hand, serve the user with satisfying results. Hence, a user-centric design plays an important role. Therefore, we conducted a survey with 184 participants to gain insights into the usage patterns and requirements of users for *Smart City Search*.

This paper is structured as follows. In Section 1 the design of the questionnaire is presented. Section 2 analyses the results and draws conclusions for each topical scope of questions. Related work is discussed in Section 3. Finally, Section 4 concludes the paper.

1. Survey Design

We have created a questionnaire[3] with a total of 44 questions, where, if not specifically mentioned, a single answer is allowed. The questions are divided into the five parts *GeneralQ*, *CreativeQ*, *GeneralUsageQ*, *ScenarioUsageQ*, and *ResultQ*. *GeneralQ* consists of questions dealing with the age, gender and prior knowledge about IT, Google services, and Google search functions. We used Google as a representative because it is the best known search engine provider. The block ends with the question if the participant heard the term "Internet of Things" before and if so, he should give a short explanation of the IoT. Answers from this block allow classification of the participating users. Afterwards, a description of the IoT is given so that the user understands the upcoming questions regarding search in the IoT. In *CreativeQ*, the participant should write down IoT search scenarios he can think of. In the other three questions, the participant is asked to formulate queries for three different scenarios. This should give an insight if users will use a specific syntax, a keyword-based query or natural language. The question block *GeneralUsageQ* asks for general usage patterns of IoT search, i.e., which devices and where to use IoT search as well as which personal data the user would reveal to improve the results. In the second to last part, *ScenarioUsageQ*, we introduce eleven search scenarios (see Table 1) and ask for the usage patterns for these scenarios in more detail, i.e., with which devices, when, how often, where, and in which context the participant would search for them. The questions of *GeneralUsageQ* and *ScenarioUsageQ* should clarify similarities and variations in usage patterns of different scenarios. In the end, the question block *ResultQ* asks the user how results should be displayed. We wanted to know, how many results to display on the first result page, which information to show, which accuracy the sensor state should have, and how to sort the found results. In this way, we can determine how result pages should be designed to satisfy the users' requirements.

2. Results of the Survey

The survey was online accessible for seven weeks in the beginning of 2013. It was mainly announced in mailing lists of the University of Lübeck which is associated with an uni-

[3] http://www.iti.uni-luebeck.de/fileadmin/website/Mitarbeiter/Mietz/survey.pdf

Table 1. Scenarios used in user survey.

Scenario key	Description
Bike	Search for a free bike at bike rental stations
Car	Search for a free car sharing vehicle
Taxi	Search for a free taxi
Parking	Search for a free parking spot
Table	Search for a free table in restaurants
Waiting	Search for short waiting times at fast food restaurants
Water	Search for water temperature for swimming possibilities
Item	Search for a lost item (with built-in position sensor)
Transportation	Search for position of public transportation (bus/train)
Streets	Search for jam-free routes
Place	Search for sunny, quiet places

versity clinic. The university is rather small and only offers studies related to medicine and computer science.

Standard tools (Excel) have been used to analyze the raw data. All results in plots are automatically aggregated from that except Table 2 which has been analyzed manually by grouping similar answers.

Finally, 184 people completed the survey. The gender distribution can be seen in Figure 1. The number of male and female participants in total as well as in most age groups is balanced. Due to the advertisement scheme, the age group of 20-29 years is, with 78.26%, overrepresented. This hypothesis is also supported by looking at the distribution of the profession: most participants are students (61.57%), researchers (9.26%), or persons employed in the medicine sector (7.41%). Hence, on the one hand the results are not representative for web users in terms of age and education, but on the other side the IT-related knowledge of the participants is still wide-ranging from non-expert web users to studied computer scientists working and doing research in the field of the IoT. 22.04% heard the term ”Internet of Things” before and had a vague or clear idea of it. Nearly all participants knew Google web search (98.91%), maps (100%), and the routing function of the latter (78.38%). However, only 32.61% respectively 48.37% knew Google's location-based services Latitude and Places. And with decreasing number of participants knowing a service, the number of participants using it even more decreases. In the end, the results, although not representative, are interesting and give good hints on what to keep in mind when developing sensor search systems. In the following sections, we discuss the results of the questions related to IoT search related questions in detail.

2.1. How Do Participants Envisage IoT Search? – Essay Questions

The first part of IoT search related questions, *CreativeQ*, consisted of four essay questions and called for the creativity of the participants. First, users were asked to think about scenarios in a Smart City they would search for real-world entities. Up to this point of the survey only one concrete scenario, searching for free tables in restaurants, was mentioned to not bias the answers. Nearly 70% of the participants gave an answer mentioning a total of 297 scenarios from which 55 were distinct. One can categorize these scenarios into four different classes. The most frequently mentioned scenarios along with

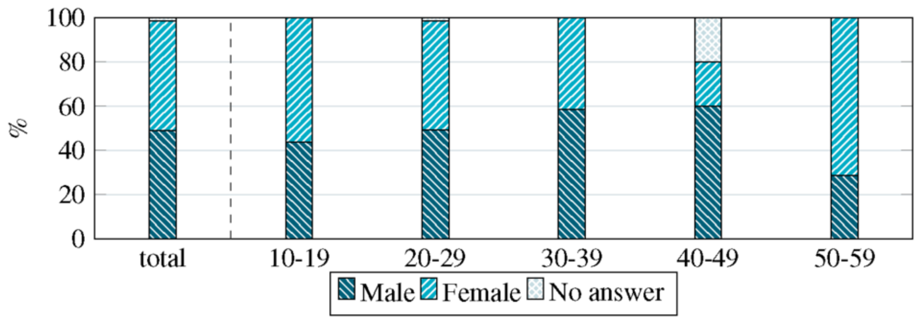

Figure 1. Total and age group divided distribution of gender.

the classes are given in no specific order in Table 2. Some scenarios can be seen in different ways and thus, categorized differently. For example, a search for car sharing vehicle can mean to find the number of available ones or the position of the nearest free one.

It can be seen that the given scenarios cover most areas of live from daily activities (shopping) and household (home appliances, heating) over learning (library) and leisure activities (sports, relaxing, culture) to social interactions (pet, friends, party). Only work is not explicitly present which can be explained by the occupation of the participants.

Table 2. Search scenarios proposed by the participants.

waiting time	**availability**	**location**	**state**
checkout lines	restaurant table	lost item	heating
attraction in amusement parc	cinema / theatre / concert ticket	next public transportation	home applicances
authority	parking spot	parcel tracking	temperature/weather
hair dresser	product in markets	car-sharing vehicle	light
doctor	seat in library	free bench	fuel price
service company	seat in public transportation	waves for surfing	water demand of plants
	book in library	friend	mood in bars
	car-sharing vehicle	pet	fill level of paper bank

The second part of this question block dealt with expressing queries to search for a *free table in a restaurant*, a *free table in a warm restaurant*, and a *free table in a restaurant with a free parking spot nearby*. The analysis shows that among the 90% of the participants who answered, most would express their request like they would do it nowadays with a web search engine, i.e., in a keyword-based fashion. However, it is also evident that many participants would like to have an interface supporting natural language querying. Finally, a small fraction of participants used a self-developed or known syntax such as SQL. The average/median number of used words for the three queries were 4.23/3, 5.93/5, and 6.11/5, respectively. This demonstrates that despite the growing complexity of the scenarios the length of queries for them is only increasing slowly. So, the total number of used words per query is low. Thus, understanding only from a query what real-world entities and states a user is looking for is a tough task.

Conclusion The results show that there is a strong need for natural language processing and understanding of the semantic meaning of a users' query because users often pose short queries with few words and without any specific syntax. To support the process of inferring the correct semantics from keyword-based or natural queries, it might be helpful to include several other aspects into the query resolution. For example, the location, time, context, and other personal information can give good indications on what the user intends to find. In the next section we investigate the characteristics of these attributes.

2.2. How Would Participants Use IoT Search? – General Search Usage Patterns

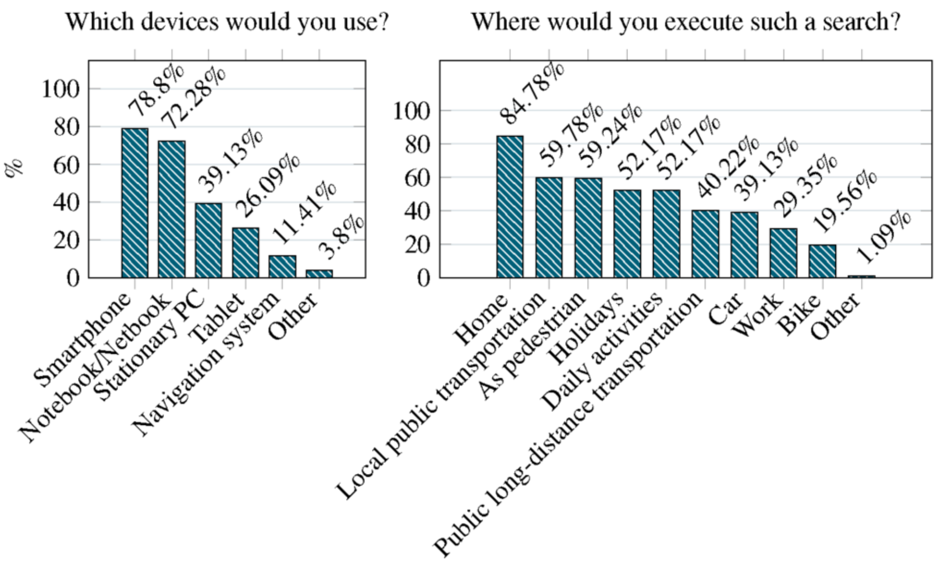

Figure 2. Results for question about general search behavior.

This part only contains questions where multiple answers are allowed and is about usage patterns of IoT search. It started by asking with which device and where the participant could imagine to pose a search query. The result plots are given in Figure 2. One can see that around 79% would use a smartphone. However, also mostly stationary devices such as notebooks and desktop computers are considered as an appropriate device to perform a search. Later on, the same question is asked for specific scenarios which will show that device choice is strongly dependent on the scenario. Although *Home* has the highest total percentage, the sum of all mobile usage scenarios exceeds home usage. The percentages for different types of mobile usage scenarios such as pedestrian or in public transportation are quite similar.

The third question is concerned with privacy issues. We wanted to know which personal information a user would pass on to the search engine to improve the search results. Especially when it comes to real-world entity search, information such as the position of the user, the preferences, or the religion can help to find the nearest or most appropriate restaurant, e.g., a vegetarian or kosher one. Figure 3 reveals that nearly $^2/_3$ have no problems to expose their position. Also basic information such as gender, age,

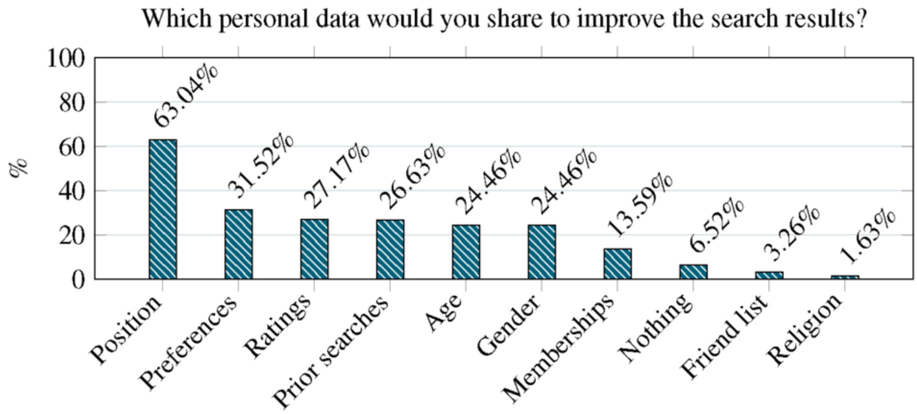

Figure 3. Results for question about sharing personal data.

and preferences would each be given away by more than 24%. On the other side, nearly nobody wants to bare his religion or friend list. 6.52% even want to reveal anything.

Conclusion Because search is used with several device classes, it is important to design search interfaces that are specifically tailored to the characteristics of each device. This includes, among other aspects, the screen resolution, available bandwidth, and supported user interactions. On a big screen all results should be displayed on a map in addition to a list view. In contrast, a low resolution screen should only show one of the views at a time but allow switching between them. Additionally, different contexts should also be considered at design time. When driving a car or bike, it would be useful to allow free-handed, distraction-free, speech-based query input.

Privacy issues are a tremendous challenge in the IoT when millions of sensors could potentially track people continuously. So, the user should have at least full control over his user search profile, i.e., what information he wants to reveal to improve his search results to build up trust in the Smart City.

2.3. How Would Participants Use IoT Search? – Usage Patterns of Search Scenarios

After asking about the general usage patterns of IoT search in *GeneralUsageQ*, *ScenarioUsageQ* aims to gain insight into the usage patterns for the scenarios listed in Table 1. Again, as in the previous block, for all questions multiple answer are allowed.

Just as in *GeneralUsageQ*, users would mainly use smartphones. Also, for other device classes the ratios are comparable to the one in *GeneralUsageQ*. Only the *Parking* and *Streets* scenario would be used by nearly $^{1}/_{3}$ of the participants with a car navigation system. These two scenarios are only outnumbered by smartphone usage.

Instead of detailed answer options about the usage place as in *GeneralUsageQ*, we broke the question down and asked where (home, on the move, at work) and in which context (spare time, journey, work, daily activities, sport) the search scenarios would be applied. Figure 4 shows that five of the scenarios (*Car*, *Table*, *Item*, *Water*, *Place*) are more frequently used from home. Searching at work occurs rarely. As already pointed out earlier, this might be due to the small number of participating employees. This conjecture is supported by looking at the search context (Fig. 5). Leisure activities and journeys/trips

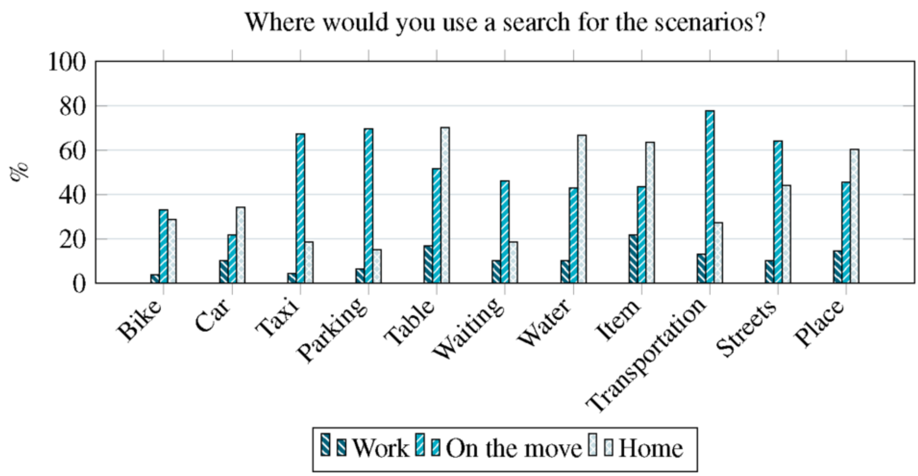

Figure 4. Results for question about usage place.

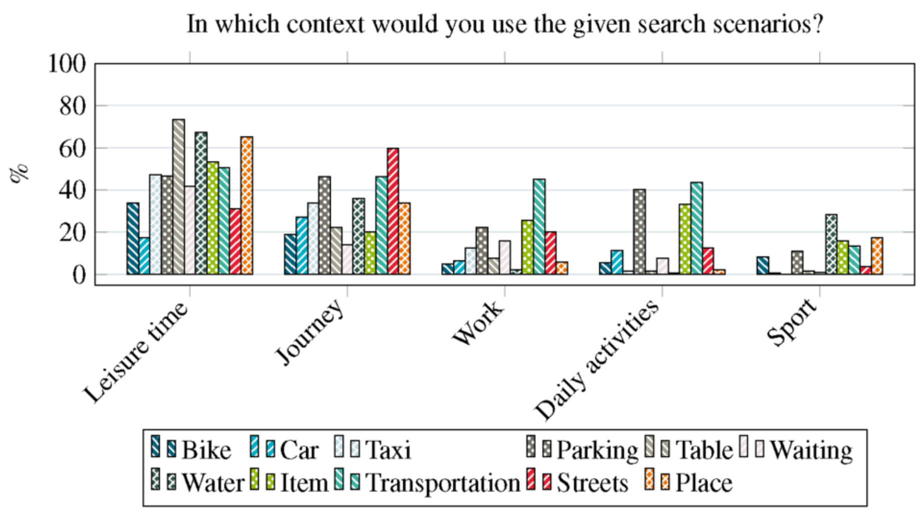

Figure 5. Result for question about search context.

are the most frequent situations to perform a search, each being more than twice as frequent as in other contexts.

The last two questions are related to time. First, we wanted to know at which times of the day search scenarios are applied (Fig. 6) and second, how often scenarios are performed at all (Fig. 7). For search scenarios such as *Item* and *Transportation*, the frequency is approximately uniformly distributed over the day while, e.g., *Taxi* and *Table* are at least three times more frequent in the evening than at any other time of the day. The frequency distribution in Figure 7 shows that many scenarios are never used. However, we believe that people may realize the comfort of such search scenarios, accept the technology incrementally, and adapt their behavior as with other technical innovations such as location-based services, Facebook, or the personal computer.

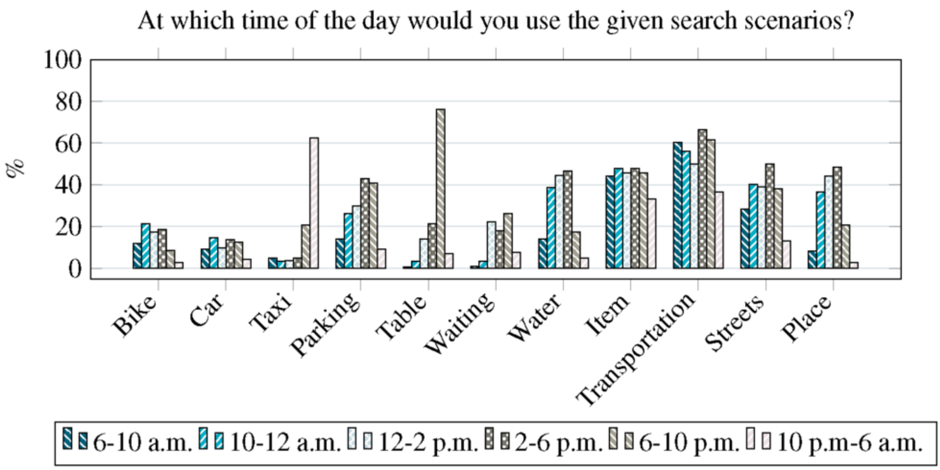

Figure 6. Result for question about the usage time.

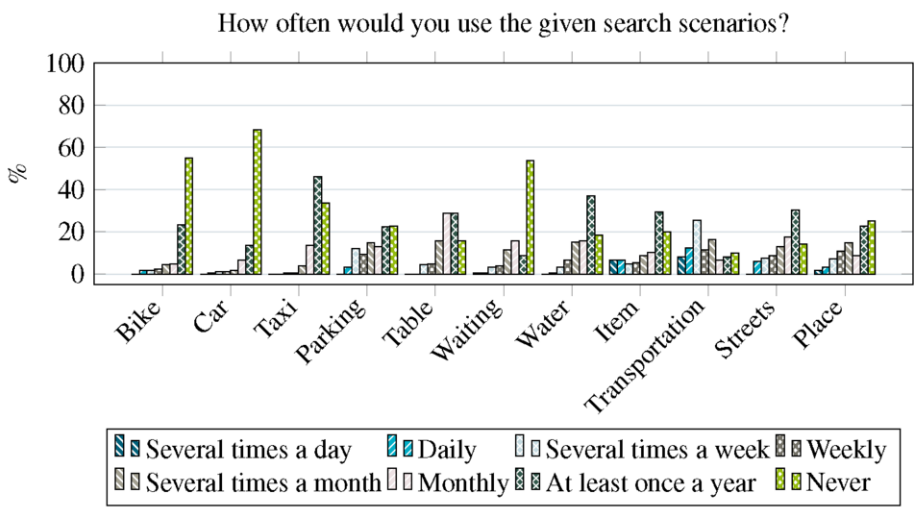

Figure 7. Result for question about the usage frequency.

Conclusion The results allow the conclusion that there is no single device or context which is mainly used but is strongly dependent on the search scenario. Thus, as mentioned in section 2.2, several characteristics of different devices and contexts need to be taken into account when designing a *Smart City Search* service.

Also, one can exploit the changing workload of sensors in the different scenarios. Take the *Table* scenario where nearly 80% of search requests are in the evening and assume there are only few state changes. Instead of contacting the sensor for each new user query, the sensor should push his state changes to the server. Contrary, if there a lot of state changes nobody or only a few users are interested in, the state can be pulled by the server whenever needed. Thus, the number of messages between sensor and server can be reduces, resulting in lower resource consumption.

2.4. How Do Participants Expect Their Results to be Presented? – Search Result Presentation

The last question block of the survey, *ResultQ*, is about how search results should be presented to the user. This includes the number of results to display, the information to show, the sorting of results, and the granularity of the state of the entity.

When it comes to the number of results on the first result page, most of the participants (77.14%) want to have ten while only 8.57% want 100, 7.43% three, and 6.86% all results. This outcome supports the approach taken by several existing sensor search systems and techniques such as Sensor Ranking [4] which try to find only a subset of matching sensors to reduce resource consumption in terms of bandwidth and energy. If more results are needed, these can be retrieved incrementally. Thus, if a user is satisfied with search results on the first page, unnecessary communication for further result pages can be avoided. Still, the question remains which information a user wants to be satisfied with the results.

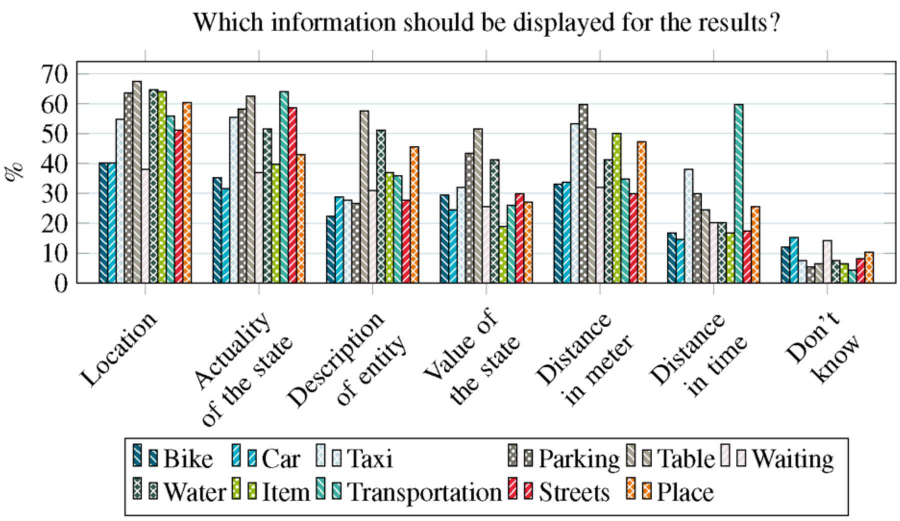

Figure 8. Results for question about the result information to display.

In Figure 8 it is shown which information users wish to see for the scenarios. Multiple answers were allowed. Although the maximum fraction of participants who request a specific piece of information is slightly above 60%, each requested piece of information is desired by at least 15% of participants. Interestingly, the *Actuality of the state*, i.e., the time since the last reading of the sensor, is in all scenarios one of the most important pieces of information. Thus, it can be assumed that users are aware that states of real-world entities can change very frequently and the older a value is, the more useless it is for the user. Consequently, it is important to have a recent sensor reading. If this is not available due to the knowledge that the state of the searched sensor is only changing rarely, an explanation of this fact should be displayed to the user.

Figure 8 shows that also the value of the state is important information for the user which leads to the question how this value should be represented. In Section 2.1 we introduced four classes of scenarios. Moreover, search results in each class share a common

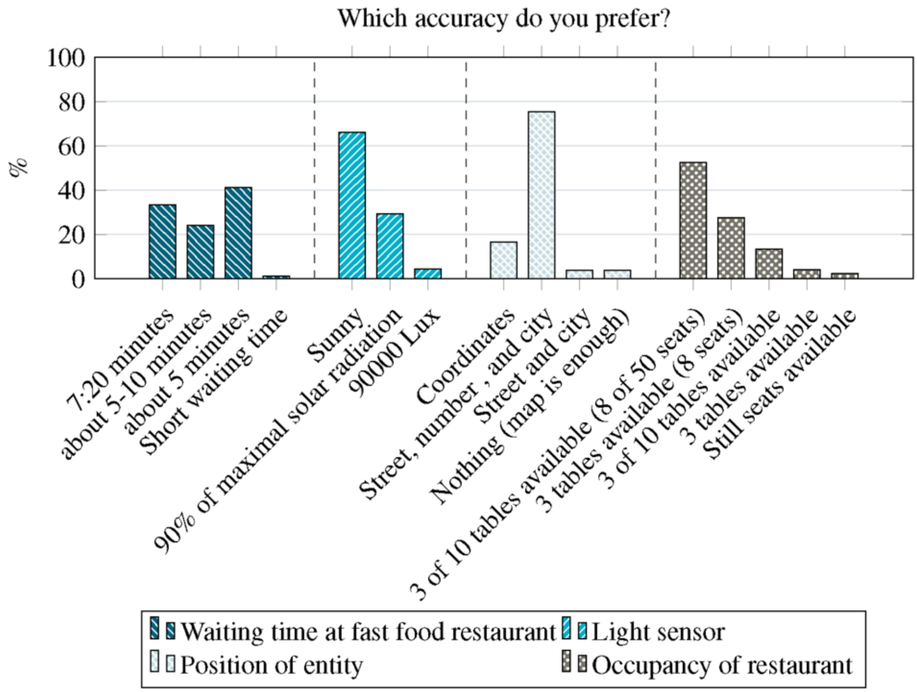

Figure 9. Results for question about the accuracy of states.

data type. The *waiting time* class has a result data type specifying time, the *availability* category has a Boolean or numeric result data type representing the number of units left, the *location* one has an address or relative or abstract position result data type, and the *state* class can has text or numeric result data type. We asked the participants which accuracy they would like to have displayed for different kinds of sensors respectively search scenario classes. The results can be seen in the plots in Figure 9. While the participants prefer an abstract state for a light sensor, they want the most detailed information for the occupancy of the restaurant. For the waiting time there is no favorite granularity. Finally, the position of an entity should be, in addition to a map, given as a fully qualified address and not as abstract coordinates in terms of Latitude and Longitude.

The sorting criteria for the scenarios was the last, multiple answer, question of this block. Figure 10 shows the results. The most important factor for users is the distance, either in meters or in time units. Especially the time until the next public transportation arrives sticks out. This can be explained by the annoyance of students about the local public transportation which is quite unreliable. The ratings of the searched entity or the neighborhood it is located in are not primary measures. Similar to the question about which information should be displayed the freshness of the state is also an important factor for the user.

Conclusion By incrementally searching for results instead of fetching all at once, the consumed resources in the backend, especially on sensor level, can be reduced, supporting approaches such as Sensor Ranking [4].

The desired accuracy for states shows that one needs flexible mechanisms in the backend to truncate or aggregate sensor values or to generate high-level states from raw

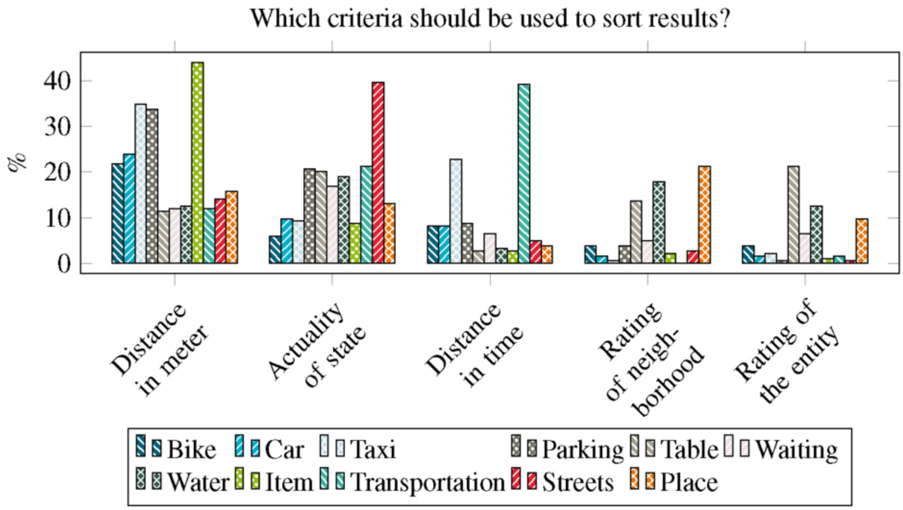

Figure 10. Results for question about result sorting criteria.

readings. This can in turn reduce the required resources because updates of high-level states are less frequent than changes in underlying sensor values as we have shown in [5].

Finally, the freshness of the state is one of the most important factors. Hence, it should be taken advantage of caching and algorithms which adaptively switch between pull- and push-based state dissemination. Users should be informed not only about the time since the last update, but also about the confidence in a displayed sensor state.

3. Related Work

The myGander [6] system is a search engine for mobile phones which allows to search for real-world entities such as a thrill ride in a amusement park. A query can have constrains like ”waiting time $<$ 20 minutes” and relevance metrics such as wait time and distance. A user study with twelve participants revealed that most users liked a real-time search. However, the small number of participants and the restriction of examining the satisfaction with the existing system does not allow to draw general conclusions for the design of a sensor search architecture.

Guinard et al. [7] also carried out a user study, but form the developer’s side. They asked 69 novice software programmer to implement a mobile phone application which retrieves sensor data both via a RESTful interface and a WS-* service. On the one hand, developers preferred REST because it was easier to learn. However, for applications with QoS and security requirements WS-* services favored. Anyhow, our user study focuses on the user perspective rather than on the developer perspective.

The IoT comic book [3], which partially inspired our search scenarios, describes different possible future Smart City scenarios in an understandable, enjoyable way. Because it has been designed for a broad non-professional audience, it does not give any development hints or research directions.

4. Conclusion and Future Work

The Internet of Things slowly becomes reality. This trend enables new services and applications. We argued that one important service will be *Smart City Search* where users can find real-world entities with a certain state. To understand how people would use such a service, we conducted a survey with over 40 questions, discussed the results and gave advise for the design of the user interface and backend, i.e. the server and sensor software, of a sensor search architecture.

There is a need for a flexible, open system to allow applications for several device classes. Furthermore, the backend must support mechanisms to conserve energy in resource-limited devices such as smartphones and wireless sensor nodes. Finally, the complete development process should be user-centric, i.e, one has to carefully deploy, explain, and give insights into the working of sensor systems in a way that allows users to get familiar with these new technologies to build up trust in them. Otherwise, with closed, black-boxed systems, users will be anxious and suspicious of sensors tracking their activities.

Based on our previous work and the outcome of this survey we intend to build a sensor search system with interfaces optimized for mobile apps and powerful web apps. Additionally, an API to the search service will allow to build other applications on top. In the backend we will exploit Sensor Ranking and adaptive algorithms to reduce resource consumption on sensor level.

Acknowledgment

We would like to thank all participants of the survey. Special thanks goes to Lukas Ruge and Nina Hoppen for testing the survey and giving helpful feedback.

References

[1] A. Caragliu, C. Del Bo, and P. Nijkamp, “Smart cities in europe, 2009,” in *3rd Central European Conference in Regional Science*.

[2] L. Ruge, B. Altakrouri, and A. Schrader, “Soundofthecity - continuous noise monitoring for a healthy city,” in *5th International Workshop on Smart Environments and Ambient Intelligence (SENAmI 2013) at IEEE International Conference on Pervasive Computing and Communication (PerCom 2013)*, (San Diego, California, USA), March 18-22 2013.

[3] M. Presser, ed., *Inspiring the Internet of Things!* Alexandra Institute, 2011. http://www.alexandra.dk/uk/services/Publications/Documents/IoT_Comic_Book.pdf.

[4] B. M. Elahi, K. Römer, B. Ostermaier, M. Fahrmair, and W. Kellerer, “Sensor ranking: A primitive for efficient content-based sensor search,” in *IPSN 2009*, IEEE Computer Society, 2009.

[5] R. Mietz, S. Groppe, K. Römer, and D. Pfisterer, “Semantic models for scalable search in the internet of things,” *Journal of Sensor and Actuator Networks*, vol. 2, no. 2, pp. 172–195, 2013.

[6] J. Michel, C. Julien, J. Payton, and G.-C. Roman, “The gander search engine for personalized networked spaces,” tech. rep., Technical Report TR-ARiSE-2011-008, The University of Texas at Austin, 2011.

[7] D. Guinard, I. Ion, and S. Mayer, “In search of an internet of things service architecture: Rest or ws-*? a developers perspective,” in *Mobile and Ubiquitous Systems: Computing, Networking, and Services* (A. Puiatti and T. Gu, eds.), vol. 104 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 326–337, Springer Berlin Heidelberg, 2012.