DESIGNING AN INTERNET OF THINGS SEARCH ENGINE;
A USER-CENTRED APPROACH

DIMITRIOS MARKOGIANNIS

This dissertation was submitted in part fulfilment of requirements for the degree of

MSc Information Management

DEPT. OF COMPUTER AND INFORMATION SCIENCES

UNIVERSITY OF STRATHCLYDE

AUGUST 2017

DECLARATION

This dissertation is submitted in part fulfilment of the requirements for the degree of MSc of the University of Strathclyde.

I declare that this dissertation embodies the results of my own work and that it has been composed by myself. Following normal academic conventions, I have made due acknowledgement to the work of others.

I declare that I have sought, and received, ethics approval via the Departmental Ethics Committee as appropriate to my research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to provide copies of the dissertation, at cost, to those who may in the future request a copy of the dissertation for private study or research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to place a copy of the dissertation in a publicly available archive.

(please tick) Yes [ ] No [ ]

I declare that the word count for this dissertation (excluding title page, declaration, abstract, acknowledgements, table of contents, list of illustrations, references and appendices is.

I confirm that I wish this to be assessed as a Type 1 2 3 4 5 Dissertation (please circle)

Signature:

Date:

Abstract

The paradigm of Internet of Things (IoT) and the rapidly expanding deployment of smart devices are expected to revolutionise the way people live. Along with the great potentials, Internet of Things faces major challenges. One of the challenges that has been identified in the literature is the capability of harnessing the large volume of data that is being generated by the smart devices. In order to identify, collect, analyse and present the data that is relevant to the users, the implementation of an IoT search engine is considered critical.

This research aims to explore the user adoption of the IoT and their requirements in order to propose a basis on which a user-friendly IoT search engine should be implemented. Hence, this research consisted of exploring the extent to which the IoT has been integrated in peoples live, discovering what information would the users like to retrieve from the smart devices and finally to identify the important features that an IoT search engine should have in order to serve the users' needs.

It was discovered that people have adopted the IoT concept to a great extent, however, the usability of the current IoT search engines (Shodan and Thingful) is not yet sufficiently developed for the users. In addition, it was revealed that the users consider as the most important the data from smart home cameras and traffic cameras. The data from water quality sensors and room conditions sensors have also been identified as very important from the users. Finally, the capability of retrieving the data immediately after the query, the aggregated search, the facets and the text snippets have been identified as the user-preferable features for an IoT search engine.

Acknowledgements

The author wishes to thank all the students who voluntarily participated in the focus groups as well as all the users who provided their feedback in the online questionnaire. Finally, the author wishes to thank the supervisor of this research project, Dr Martin Halvey, for his guidance and support.

Contents

List of tables

List of figures

List of abbreviations

IIR: Interactive Information Retrieval
IoT: Internet of Things
IR: Information Retrieval
RFID: Radio-Frequency Identification

1. Introduction

Internet of Things (IoT) is a relatively new paradigm with the potential to completely transform the way people leave. In August 2015, IoT concept was placed in the peak of Gartner's Hype Cycle for emerging technologies with the expectance to reach the so called "plateau of productivity" in five to ten years (Rivera and Van der Meulen, 2015). Even though the IoT paradigm and its applications are not yet fully integrated in people's everyday lives, the deployment of the supporting smart devices (also called things), is rapidly increasing. Earlier this year (January 2017), Gartner forecasted that the connected devices used worldwide, will reach the number of 8.2 billion in 2017 and up to 20.4 billion until 2020 (Van der Meulen, 2017). The IoT concept covers a wide range of applications extending from simple RFID trackers for personal use to a complex network of sensors for predicting and warning about disasters in the mining industry (Li Da, Wu and Shancang, 2014). The Internet of things term was firstly used in 1999 by Kevin Ashton, as a title for a presentation at Procter & Gamble (Bandyopadhyay and Sen, 2011; Ashton, 2011). Nevertheless, almost 20 years after the concept of IoT was mentioned for the first time, the academic community has not yet concluded to a single unified definition. On the contrary, by reviewing the literature, various definitions can be found, each one focused on the purpose of the research. In this research, IoT is defined as the concept in which various and diverse devices, called smart things or objects, can be controlled, interact and exchange information with each other or their environment, via the network to which they are connected. IoT encompasses the smart devices which are used to just sense and/or monitor the surrounding physical environment as well as the devices that are actuators or can signal another smart device to operate. As Cavalcante *et al*. (2016) assert, a network of smart objects is capable of assisting the users and providing valuable information without the need of human intervention. In 2008, the US National Intelligence Council estimated that by 2025 everyday objects such as food packages, furniture and paper documents will nest internet nodes, and that the demand in combination with the technological advances "could drive widespread diffusion of an Internet of Things (IoT) that could, like the present Internet, contribute invaluably to economic development and military capability" (National Intelligence Council, 2008). As it is expected to revolutionise the way people live, like the internet did, the infrastructure might fall short of supporting such a technological breakthrough. The tremendous increase of heterogeneous devices leads to a massive volume of data being transmitted. Consequently, the ability to identify, collect and harness the vast amount of data, poses a major challenge for the deployment of IoT (Cavalcante *et al.*, 2016). As a result, the deployment of an IoT search engine is considered crucial in order to overcome the latter challenge. Similarly to the internet, where all the available sites, information and services would be of no value to the end user if they were not discoverable and accessible, a search engine crawling the internet to locate the existing smart devices, collect and analyse the data, would increase their usability. Searching, retrieving and analysing the information and data in the internet is very common with the utilisation of the traditional search engines such as Google and Bing. However, the number of equivalent applications for supporting the retrieval and analysis of the heterogeneous and sparse data generated from smart devices, is limited to two; Thingful and Shodan. Despite the fact that a lot of research has been conducted and many different IoT search engines have been proposed in the literature, only Thingful and Shodan have been deployed for public or commercial use, utilizing real-world data (Shemshadi *et al.*, 2016). Even though Thingful is the application closest to the traditional search engines, further progress is needed to overwhelm the limitations that occur due to the lack of publicly available data and a user-friendly interface. (Shemshadi, Sheng and Qin, 2017). Shodan was primarily designed for hackers, providing access even to password protected smart devices, however, now is specifically designed for commercial use. Companies in different industries can use Shodan either to detect vulnerabilities in their security, or to exploit IoT data for strategic purposes (*Shodan*, 2017); nevertheless, Shodan does not provide any support to its web interface. Subsequently, it is obvious that even though the IoT paradigm rapidly expands and the volume of data transmitted over the internet exponentially grows, there is very little progress in regards to the IoT search engines. Noticeably, further research is needed to discover how the IoT search engines can be improved in order to underpin the rapid adoption of Internet of Things as well as to be incorporated in people's everyday lives. Several questions rise on the topic both from technical and user perspectives. The current research, focuses on the user perspective and aims to examine at what degree the users have adopted the IoT paradigm, as well as their requirements. In addition, it aims to provide a baseline for future research related to the user-centred perspective of the IoT search engines. In order to accomplish the research's goals the following questions must be answered:

1. To what extent have the users adopted the Internet of Things paradigm?
2. What data is considered most valuable for the users?
3. What features people would want in an IoT search engine to be more usable?

This research focused on examining what people expect from an IoT search engine and which characteristics would lead to greater user satisfaction and consequently to greater adoption of the IoT search engine. Specifically, in this dissertation it is investigated what people want to find using an IoT search engine, how they want the retrieved information to be displayed and finally what features they prefer to be incorporated in an IoT search engine.

The content of this report is organised as follows: In section 2, the related work is presented. The gap in literature regarding the IoT search engines is identified and the reasons for conducting this research are highlighted. In section 3, the first part of the research is demonstrated which included the online questionnaire. In the sub-section 3.1 the methodology that was followed in each part of the questionnaire (3.1.1 – 3.1.5) is described, along with the justification of every decision made. In the sub-section 3.2 the results of the online questionnaire are presented as well as the analysis that was conducted to reach to conclusions. In section 4, the second part of the research is demonstrated, which consisted of gathering the users' requirements via the method of focus groups. In sub-section 4.1, the methodology that was followed is presented. The sub-divisions of the latter (4.1.1 - 4.1.5) contain detailed justification for every decision made in each part of the focus group respectively. In section 4.2 the results of the focus group are presented along with the corresponding analysis that was conducted in each part (4.2.1 – 4.2.3). Finally, in section 5 this research concludes about the findings and provides recommendations for further research.

2. Literature review

As proves from the literature review below, the user perspective has been thoroughly investigated for different types of Information Retrieval (IR) systems. With the novel paradigm of IoT rapidly expanding, emerges the need for a search engine to support it. Several studies have been conducted to propose the technical frameworks which would lead to the successful implementation of an IoT search engine. In addition, several other researches have pointed out the challenges, including the legal and ethical challenges, that rise with the deployment of different IoT applications. However, the user factor has not been taken under consideration to the same extent. The user acceptance of IoT and the user's requirements for an IoT search engine remain unexplored. In this section the literature about the IoT and the IoT search engines is being presented. In addition, several other studies are presented which explore different IR systems from a user-centred approach. Moreover, it is identified that current literature for IoT search engines, focuses more on the technical aspects leaving the user factor unexplored. Finally, it is explained that the aforementioned gap in literature, acts as a motivation for conducting this research.

As IoT is a relatively new paradigm, the related research made so far is very limited compared to other concepts in computer science; nonetheless, the popularity of IoT is rapidly increasing in the academic literature. Botta *et al (2016)* analysed the published studies and papers regarding the IoT, aiming to discover its popularity among the researched topics, and concluded that there has been a significant increase since 2008. On a similar approach, Cavalcante *et al.* (2016) also reviewed the existing literature referring to the IoT concept, but aiming to explore the topics that most of the published research was focusing on. They ascertained that most of the researches, could be classified in five dominant categories; architecture, platform, framework, middleware and challenges, with some studies assigned to more than one categories. The latter can be confirmed when focusing on the literature that refers to the IoT search engines. The concept of an information retrieval system designed to support the IoT has been explored in the academic literature, although not as extensively as other topics. Internet of Things has been studied to a great extent, but as Ding, Cheng and Yang (2014) point out, the techniques for collecting and managing the growing volume of data transmitted by smart objects, need more research. Hence, the authors of the latter study, have concluded that an IoT search engine should ideally support the following:

1. Real-time retrieval of sampling data generated from smart devices.
2. Spatial conditions of the search.
3. Value-based search conditions (instead of keyword based) as the data is numerical.
4. Past and present data for a more accurate analysis of the data.

Ostermaier *et al*. (2010) also acknowledge that traditional search engines and IR techniques are not sufficient to tackle the challenges rising from the diversification and the volume of data generated and transmitted by smart devices. Reviewing the literature, an adequate number of published researches can be found about IoT search engines, however, most of them approach the topic from a technical perspective. One of the first systems that can be considered IoT search engines was called Max and aimed to assist users to search and locate objects in the physical world (Yap,Srinivasan and Motani, 2008). However, it lacked the ability to retrieve any other information about the object apart from its location. Later on, Wang, Tan and Li (2008) *w*ere the first to propose a more complete IR system, specifically designed to utilise a network of sensors in order to assist people to search objects of the nearby physical environment. The rapid changes of the data objects retrieved from the sensors and the differences with the data objects in databases were taken under consideration before the design of the proposed IR system. In their paper, are demonstrated the system components and architecture, the processing techniques as well as the technique followed for the needed communication to be established. On a similar approach, another design called 'Microsearch', has been proposed to assist users in their search for information from devices used in pervasive computing (Tan *et al.*, 2010). Unlike the previous designs and the search engines as known, Microsearch was designed to index the information inside the embedded devices so that common IR systems could identify the devices and accurately decide about their relevance with the user's query. As the creators assert, Microsearch was designed as an assistive component that could be integrated in other physical world search engines. The aforementioned research paper focuses on the architectural perspective of the proposed solution and provides detailed description of the methods and techniques used to achieve the expected outcome. Another experimental design of an IoT search engine that can be found in the literature is 'Thingseek' (Shemshadi,Sheng and Qin, 2016). Thingseek is a proposed framework for a search engine which includes the user interface in its design, as well as the way it handles the keywords and human interaction. Despite the numerous IoT search engines that are being proposed in experimental level, there are only two that are already deployed and addressing to real-world situations; Thingful and Shodan. Due to the inaccessibility to the available data, Shemshadi *et al. (2016),* point out that "Thingful is still limited and significant progress is needed to deploy in real-world or large-scale data". Shodan on the other hand, as it was used primarily as a hacking tool, gives the ability to sophisticated users to hack into password protected devices. The aforementioned ability of Shodan, has raised a lot of awareness both in the media and the academic community. As a result, a plethora of studies focus on Shodan, exploring the ethicality of its existence and potential misuse as well as its capabilities as an IoT search engine. However, Shodan's capabilities can be used in exchange of a fee and the interface addresses to technically knowledgeable users. For that reasons, Shodan is mainly used by some industries to identify weaknesses in the security of their equipment connected to the internet. Bodenheim *et al.* (2014) state that "Shodan provides attackers with a powerful reconnaissance tool for targeting industrial control systems". Acknowledging the above, they have evaluated Shodan's indexing and query ability and they proposed a strategy to limit the exposure of the industrial control systems to Shodan queries. Based on Shodan's capabilities, another tool for vulnerability assessment has been proposed called ShoVAT (Genge and Enăchescu, 2016). Recognising the importance of the industrial IoT applications, Lunardi *et al.* (2015) have presented another software framework, called COBASEN, consisting of "a Context Module and a Search Engine", which can promote the development of industrial IoT applications. As it can be observed, despite the fact that a variety of researches have been made towards the implementation of a search engine for collecting, analysing and differentiating the data generated from the smart devices, little effort has been made to capture the users' preferences on such a tool. Shemshadi *et al*. (2016) discovered that "public interest towards IoT with its most popular abbreviations has been steadily increasing over the past few years" and approached the IoT topic from a more user-centred perspective. They analysed frequent queries and keywords online aiming to identify user interests related to the Internet of Things and they summarised their findings in a list with the most popular keywords and their categories as can be seen in table 1.

Arguably, the participation of the users in the concept of IoT is a "key requirement that needs to be considered in the Internet of Things" (Uckelmann, Harrison and Michahelles, 2011). All things considered, there is a noticeable gap in literature regarding the requirements of the users from an IoT search engine, how do users assess the retrieved information and what features they would like to be integrated in the IoT search engines. However, the same gap is not observed for more common

Information Retrieval systems for which, user perspective as well as the human interaction have been studied thoroughly.

	keyword	frequency	category	%
1	air quality	71,700	environment	61.7
2	sensor	3,348	misc.	2.8
3	ship	1,851	transport	1.6
4	radiation	1,825	environment	1.5
5	earthquake	1,601	environment	1.4
6	gamma	1,131	environment	1
7	weather	876	environment	0.8
8	shark	851	flora and fauna	0.7
9	temperature	581	environment	0.5
10	camera	397	home	0.3
11	car	392	transport	0.3
12	iphone	271	home	0.2
13	fridge	259	home	0.2
14	webcam	255	home	0.2
15	aircraft	247	transport	0.2
16	sharks	245	flora and fauna	0.2
17	energy	242	energy	0.2
18	food	239	home	0.2
19	netatmo	216	environment	0.2
20	coffee	177	home	0.2
21	traffic	168	transport	0.1
22	transport	166	transport	0.1
23	cars	163	transport	0.1
24	raspberry pi	159	experiment	0.1
-	other keywords	28,771	-	24.6
-	Total	116,131	-	100

Table 1 - Most popular keywords and their categories (Shemshadi et al., 2016)

For instance, acknowledging that “the importance of user consideration in the design of information retrieval systems has been recognised for a long time”, David Ellis (1989) followed a user-centred approach to the IR systems and explored the features that should be implemented taking under consideration users’ behavioural and seeking patterns. Dumais *et al. (2016)* presented a design of an IR system to assist users to retrieve information which have found in the past. After thoroughly describing the architecture and the interface of the proposed IR system, the researchers proceeded to its evaluation by gathering quantitative and qualitative data from 234 users, highly focusing on the interface and the users’ preferences. Several other studies can be found evaluating the interfaces of different IR systems designed for specific purposed. Similarly, after the implementation of a video IR system, the researchers proceeded to its evaluation by gathering feedback from actual users and investigating potential improvements (Hopfgartner *et al., 2008*). Other studies focus on the interfaces of IR at experimental level. For instance, Belkin, Marchetti and

Cool (1993), presented a design of an interface of an IR system used to retrieve bibliographic information. The designed interface, called BRAQUE, aimed to “support user interaction in an integrated fashion” considering the information seeking strategies and the search behaviours of the users. The study consisted of the framework and techniques used to retrieve the relevant data, as well as a demonstration of the user interface which assists the user to discover the information needed. Moreover, in a similar research, detailed scripts have been presented for the successful integration of the user interaction and the greater effectiveness of the IR system. As the researchers of the latter study point out:

> It is becoming increasingly evident that IR is an inherently interactive process, from a variety of points of view. This means in particular that supporting and taking advantage of the interaction of the user with the other components of the IR system is crucial for effective IR system design.(Belkin *et al.*, 1995, pg379)

In other studies which include surveys, the participants were asked to express their preference among a variety of interfaces or express their opinion about the interface of an IR system. In addition, accepting the fact that the interface of an IR system can affect the user behaviour and perception about the relevance of the results, the effectiveness of the IR systems is also studied by evaluating the respective interface. All the studies aim to gather user requirements with the goal to achieve an improved user interaction and experience. In that respect, Hu, Ma and Chau (1999), conducted an experiment in which 601 participants were asked to evaluate different interfaces based on the perceived relevance of the information retrieved. The final evaluation and comparison of the designed interfaces was conducted by analysing the data following the ANOVA approach. They concluded that “interface design may have a significant effect on system–user concept communication and that graphical user interface may be more effective in supporting such communication than a list-based design”. A complementary study has been conducted, in which the IR system that was investigated was Google; the most popular real-world search engine. The research involved 26 users who were tasked to perform four searches with different queries and then they were asked to evaluate the relevance of the documents retrieved as well as to express their satisfaction about the accuracy, coverage and ranking of the results. (Al-Maskari, Sanderson and Clough, 2007). Petrelli *et al.* (2004) followed a different strategy for the implementation of an IR system, called ‘Clarity’, which purpose was to find documents in different languages. The main difference with the previous studies is that the users were involved from the initiation of the project until its final development and evaluations. The process included use case scenarios from the beginning in order to define the actual needs of the users as well as during the design and implementation phases to gather users’ requirements and adapt the IR system accordingly. Furthermore, the interface of Clarity was designed after a number of mock drawings were created in co-operation with the users. Subsequently, a notable portion of the literature mentions the importance of user interaction in the evaluation of the Information Retrieval systems. The Interactive Information Retrieval (IIR) system is defined as the system which encompasses the user's needs and behaviour and enables the user to dynamically interact with the IR system to complete a search task. (Borlund, 2009). Based on the latter definition the model for the appropriate evaluation can be divided in three parts:

1. A set of components which aims at ensuring a functional, valid and realistic setting for the evaluation of IIR systems,
2. Empirically based recommendations for the application of the concept of a simulated work task situation; and
3. Alternative performance measures capable of managing non-binary based relevance assessments (Borlund, 2009, pg 29).

As part of the set of components mentioned above, is considered the engagement of users and probable stakeholders in the experimental phase of the IIR implementation (Borlund, 2000). One approach for the users' involvement is the simulated task situation in which the test subjects are provided with a search scenario and they are asked to complete a task. Borlund (2000) provides a comprehensive methodology of conducting different types of IIR system experimental evaluation based on the users' needs and behaviours.

Summing up, the evaluation of the IR systems following a user-centred approach has been studied extensively. In addition, a plethora of strategies, techniques and methodologies have been researched in order to design an IR system which enables user acceptance. Finally, the importance of the user preference and interaction in the effectiveness of the IR systems has been pointed out several times. However, when focusing on the IR systems (search engines) designed to support the IoT paradigm, it can be observed a skewness in the published studies towards the technical approach. Hence, the following questions rise from this contradicting observation.

- Why the user-centred approach has not been followed in the design and implementation of the IoT search engines?
- Why the already-deployed and experimental IoT search engines have not been evaluated according to a user-centred evaluation model?

The above questions act as motivation to the researcher of the current study in order to fill the gap in the literature, albeit partially.

3. Online Questionnaire

The first part of the research consisted of an online questionnaire aimed to answer the first and second research questions. Additionally, it was designed to extract information, such as the background of the users, to be able identify the strengths and weaknesses of this study, to reach to more accurate conclusions and recommendations as well as to provide the context of the study in detail, aiming to serve as a baseline for future researches. The online questionnaire was created using Qualtrics software tool and before its actual deployment, two users closely affiliated to the researcher ran the pilot tests. As the users had no prior knowledge or experience with the IoT concept their feedback helped the researcher to state the questions in a way understandable by unknowledgeable users. In addition, the feedback about the content and the way the questions presented were taken under consideration in making the decisions about the final form of the questionnaire. After the pilot test completion, the survey was distributed via the social media and was open for 4 weeks. A total of 69 responses were collected. The online questionnaire comprised of the following major parts:

1. Demographics and background
2. Familiarity with the IoT and IoT search engines
3. Smart devices usage
4. Importance of the data transmitted from different smart devices
5. Context awareness from the user's perspective
6. Contact details

The purpose and the methodology followed in each part are described in detail in sections 3.1.1 - 3.1.5.

3.1. Methodology

3.1.1. Demographics and background

At the beginning of the questionnaire the respondents were asked to provide their demographic characteristics and their background in information technology. The above information was considered crucial in order to be able to identify the weaknesses and strengths of the study, having the ability to explain any skewed results, make more accurate conclusions and recommendations for further research.

Firstly, the users were asked to choose their gender. Since the online survey was anonymised and there was no way to identify the respondent, unless submitting his personal details at the end of the survey, it was decided to provide four options (Male, Female, Other, Prefer not to say), in order to cover all potential choices of self-identification and privacy. The next questions, required users to state their age and then to specify their background related to information technology and computer science. Acknowledging that using radio buttons increase the validity and the rate of the responses (Couper, Traugott and Lamias, 2001), the method of letting users to input their age in number format was chosen over selecting the age group, so that the statistical analysis was easier.

3.1.2. Familiarity with the IoT and IoT search engines

Aiming to answer the first research question, the participants were asked to briefly explain in their own words what they understand to be the meaning of "Internet of things" and if they knew the existence of the IoT search engines. In the case that the response to the latter was "Yes", another set of questions appeared to capture the usage and the name of the IoT search engine as well as the user satisfaction.

3.1.3. Smart devices usage

Complementary to the questions asked in section 3.1.2, the usage of smart devices by the participants was investigated as well, aiming to capture to what extent the users have integrated smart objects in their everyday lives and answer to the first question of the research. To serve the latter purpose it was asked if they have used a smart device. In the case that the response was "Yes" an additional question appeared requesting to list the latest smart devices (up to three). The number of devices was limited for the sake of simplification in the stage of the data analysis.

3.1.4. Importance of the data transmitted from different smart devices

The fourth part of the online survey aimed to investigate how important was for the users to be able to discover and retrieve data from some specific smart devices, and answer the second research question. Likert-type scales were used ranging from "Not at all important" to "extremely important" and coded from 1 to 5 respectively. Several factors were considered to decide the way that the Likert-type scales would be presented and the devices that would be included for rating.

Firstly, it was decided that the devices presented for rating would be classified in three categories as shown below:

Home usage: camera, smart lights, smart thermostat, virtual home assistant and sensors for room conditions (temperature, humidity air quality etc.).
For environmental purposes: weather station, UV sensor, air quality sensor and water quality sensor.
Used in smart cities: traffic camera, sensor for parking availability, charging point and noise sensor.

Table 2 - Categories and smart devices

In all three categories there was the option of "other" which gave the opportunity to the respondents to mention other device of their preference and rate its importance as well. Arguably some devices can be classified in more than one categories, however, they were categorised only once to avoid confusion and achieve contingency in the responses. The literature lacks of a definitive list of the most popular devices that it could be included in the online questionnaire. Hence, to make the decision of which devices should be included, several sources reviewed including news web pages, blogs, smart device retailer's web sites and Deloitte's published Consumer Survey (Lee and Talbot, 2016) and articles (Dubbeldeman and Ward, 2015). In addition, Thingful and Shodan were used to review the availability of which smart devices could be discovered. All things considered it appeared that the most popular smart devices which were also discoverable were those listed in table 2.

Considering that when the dragging technique is used to replace the clicking method of responding, it benefits the results (Sikkel,Steenbergen and Gras, 2014), it was decided to use value sliders to rate the importance of the devices in the three categories.

3.1.5. Context awareness from the user's perspective

This survey also focused on the context-awareness of the IoT devices from the user's perspective to have a better understanding on what data is considered most valuable (research question 2). In the IoT era, numerous and diverse sensors are being deployed and connected to the Internet, which leads to a massive generation of data. However, this enormous volume of data is not possible to be collected and processed. As a result, the need to understand which data needs to be processed is rising and context-awareness becomes increasingly important (Perera *et al.*, 2014). Perera *et al.* (2014), support the importance of context-awareness for the IoT paradigm based on the fact that it gives the ability to capture the context of the data generated by the sensors, hence, interpret and harness that data easier. Studying the literature, a list of the "most popular keywords" (see table 1) can be found. However, it has not been studied what the users wanted to find by using those keywords. Assuming that the users expected to retrieve useful data from their queries, it was considered important to capture how the users interpret the data from a sensor and understand what data the search engine should retrieve. In that respect, a list of the most popular search terms was presented to the participants, along with a blank column, in which they were asked to write the expected devices in each case. The list included the first ten terms of “most popular keywords and their categories” (table 1) and excluding the terms that already referred to smart devices, as shown in appendix 3.

3.1.6. Contact details

At the end of the online survey, the users were encouraged to provide their contact details in case they wanted to voluntarily participate in one of the focus groups that would be conducted as part of the second stage of the research.

3.2. Results and analysis

3.2.1. Demographics and background

From the 69 respondents, 46 were male (66.67%), 22 were female (31.88%) and one (1.45%) chose the option ‘Prefer not to say’. The age of the respondents ranged from 22 to 61 with an average of 30.65. Most of the participants (33) had an academic (student, researcher etc.) background related to information technology or computer science while 20 of the respondents had no computer-related background at all and 15 of the participants had a professional background with IT. One participant did not provide any details.

3.2.2. Familiarity with IoT and IoT search engine

From the 69 responses, 61 included a description of what the users consider to be Internet of Things. For the analysis of the qualitative data, themes were created and the responses classified accordingly. The themes along with their description are demonstrated in table 3.

Theme	Description
General	Relevant description without containing any technical terms, examples or use cases.
Technical	Definition containing one or more of the technical terms: protocol, ip, server.
Strategic	Definition that included description of how IoT is being exploited strategically.
Example	Definitions in which examples of smart things were givev.
Generic	Generic terminology, making it impossible to decide if the user had or not knowledge of IoT.
No knowledge	Responses that clearly stated that the users did not have any knowledge of the concept, or the provided description was wrong.

Table 3 - Themes for the explanations about IoT familiarity

After defining the themes, the responses were classified in the relevant theme. The results are represented below in table 4.

Theme	Responses	Percentage
General	36	59%
Technical	4	7%
Strategic	7	11%
Example	13	21%
Generic	4	7%
No knowledge	4	7%

Table 4 - IoT familiarity themes and results

Some of the responses were classified in more than one themes. An increased difficulty was recognised in clustering the responses to either 'Generic' or 'No knowledge'. However, as the number of different ways of analysing the qualitative data equals the number of different researchers (Lacey and Luff, 2007) a different classification would be accepted as well. Using the word-cloud tool in Qualtrics as a means of qualitative analysis, it derives that the most used words in explaining what is the IoT, are "internet", device", "connect" while the words "thing", "network" and "connection" follow in the list. Finally, 55 (90%) of the users provided a fairly accurate description of the IoT and over one third (20) of the latter, used terms referring to machine-to-machine communication and data exchange. The full report of the responses is included in appendix 7.

From the responses collected about the users' understanding about the concept of IoT, it can be concluded that the respondents were knowledgeable about this paradigm.

Despite the fact that the users appeared familiar with the relatively new paradigm of IoT, the majority (69.6%) did not know about the existence of IoT search engines. 21 of the respondents stated that they knew about the existence of IoT search engines, of whom ten stated they have also used an IoT search engine. Five have used Shodan while the rest of the responses were discarded as the answers included "Google" "NA" or left blank. The four of the respondents who answered they have used Shodan, had academic background related to computer science or information technology and also provided the degree of satisfaction for the retrieved results. Two (50%) were extremely satisfied while each of the other two stated 'somewhat satisfied' and 'somewhat dissatisfied' respectively. Due to the small number of responses, further analysis was not conducted.

3.2.3. Smart device usage

In the next part of the questionnaire in which the participants asked whether they have used a smart device, the majority (40/67) has answered 'Yes' with the dominating device to be a smart thermostat or other heating control device (12 mentions). Other popular devices appear to be the IP camera (10 mentions), wearable devices (9 mentions), including smartwatch, fitbit or the term "wearable" in general, and smart lights (8 mentions). The respondents also mentioned a variety of other devices including virtual assistants (Alexa, Google home) and smart TVs. The percentages of the IT background of the respondents who have used a smart device (figure 1) is very similar to the general sample.

Consequently, it can be concluded that there is not significant correlation between how technologically sophisticated the user is, and the usage of smart devices. Additionally, it proves that most of the users have started integrating Internet of Things in their everyday life.

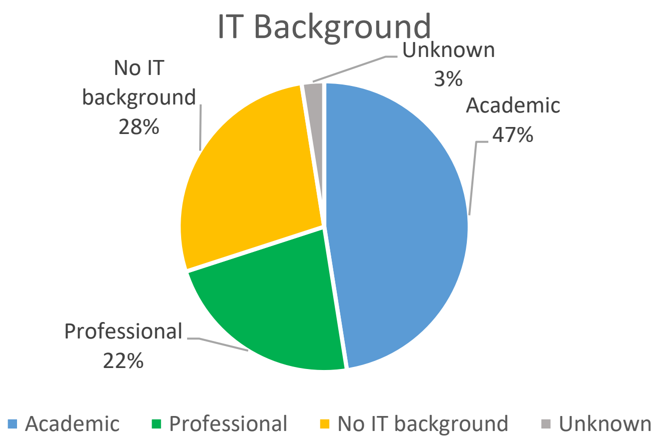


Figure 1 - IT background of people who have used smart devices

3.2.4. Importance of the data transmitted from different smart devices

The rating of the smart devices' importance, provided valuable information about the sectors and the appliances that the IoT search engines could be tested and improved, in order to encompass the users' requirements and become more useful to the public. The respondents of this online survey, assessed the importance of the devices in each category as follows:

Home	Not at all important	Slightly important	Moderately important	Very important	Extremely important	Count
Camera	1	3	10	27	19	60
Smart lights	5	17	11	17	9	59
Smart thermostat	2	12	10	25	9	58
Virtual home assistant	8	11	14	12	9	54
Sensors for room temperature, humidity, air quality, etc.	1	3	14	29	11	58
Other	5	3	9	4	5	26

Table 5 - Smart device importance – Home usage

Environmental	Not at all important	Slightly important	Moderately important	Very important	Extremely important	Count
UV sensor	3	3	10	19	20	55
Weather station	1	4	12	25	17	59
Air quality sensor	2	2	7	19	26	56
Water quality sensor	2	2	4	21	29	58
Other	5	0	11	4	4	24

Table 6 - Smart device importance – Environmental usage

Smart cities	Not at all important	Slightly important	Moderately important	Very important	Extremely important	Count
Traffic camera	2	4	8	21	21	56
Parking availability sensor	1	5	6	24	23	59
Charging point	2	4	14	24	10	54
Noise sensor	4	9	18	18	6	55
Other	7	0	13	3	1	24

Table 7 - Smart device importance – Smart cities

As this assessment was Likert-type aiming to order the importance of the smart devices and no actual magnitude between each selection was intentioned to be measured, the data considered ordinal (Göb, McCollin and Ramalhoto, 2007; Boone and Boone, 2012). However, since all the questions of this survey were optional, the number of responses for each device was different, raising the grade of difficulty in analysis, as the results are not completely comparable. For instance, the conclusion about the importance of a smart device over another, would not be valid in case the first had a greater mode but the second had greater frequency. Hence, to confirm the validity of the results, a total score, which consisted of summing all the values from 1 to 5 occurred in every instance, was calculated for each device following the approach of interval data analysis (Jamieson, 2004). From the ordinal analysis, it appears that the users considered most important to retrieve data from the air quality sensor and the water quality sensor, which had the mode value of 'extremely important'. Focusing in each category separately, the sensor for room conditions and the camera appear the two most important for home usage, having the mode value of 'very important' with a frequency of 29 and 27 respectively. In the second category, which includes smart devices used for environmental purposes, the most important appears to be the sensor for water quality, having the mode value of 'extremely important' and being ranked at that level 29 times. In the following category, which comprises devices used in smart cities, the devices which occur to be most important for the users are the sensor for parking availability and the charging point, both having the mode value of 'very important' with 24 occurrences. However, this approach fails to incorporate the frequencies of the responses for each device in each category as well as to assess the importance based on different weight for each scale.

Hence, the data analysed using the interval approach as well. To ensure the validity of the results, the data was analysed as input of a SERVQUAL survey; a popular technique to measure customer satisfaction using Likert scales. Hence, each device in this survey was considered a dimension of the SERVQUAL equivalent and the magnitude scaling technique was used. Magnitude scaling made popular by Lodge (1981) and refers to the technique of weighing and quantifying the different levels on a Likert scale (Hart, 1996). Lodge (1981) conducted an experiment, in which the participants were tasked to assign magnitudes on seven-point Likert scale ranging from 'atrocious' to 'excellent'. The results showed that there were significant differences in how the respondents quantified each grade of the scale. It was also observed that the differences were greater to the right part of the scales than in the middle, having an approximately value of 0.8 - 0.9. In a try to refine and improve the magnitude scaling, Hart (1996) took under consideration a suspected "scale effect" and after the analysis of the same sample, he assigned the magnitudes slightly differently. Nevertheless, the differences remained at 0.8 between fifth and sixth grade of the scales, and 0.9 between sixth and seventh grade. Accepting the latter as a valid measurement, in this research, the pre-determined difference of 1 (5 - 4) between 'extremely important' and 'very important' is considered accurate for calculating the importance as it is very close to 0.9. In addition, it is worth pointing out that from figure 2, figure 3 and figure 4, a skewness towards the highest levels of the scales could be observed. As a result, the differences between the lower grades of the scales were considered of less significance to calculating the final score, and the approach of all the levels having the same difference was adopted.

Summing the values of the different scales for each device, all the responses are taken under consideration weighed differently. The weights allocated as seen in table 8.

Rank	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Allocated score	1	2	3	4	5

Table 8 - Importance of smart devices conversion scale

Following the latter approach the most important device in general appears to be the camera, either for home usage (security camera, baby monitor) or used in smart cities (traffic camera), with a total sum of 240. Concluding for each category, the device considered most important for home usage is the camera with a total sum of 240 and 60 times ranked. In the category for environmental usage the results are similar with the ordinal approach, pointing as the most important device to be the sensor for water quality, having a total of 240 and ranked 58 times. In the last category, the result differ from the previous analysis, showing the traffic camera as the most important with a sum of 240 and being ranked 59 times in total. To increase the accuracy of the conclusions following this approach, the mean value as well as the standard deviation were calculated. As it can be observed, the mean values are in accordance with the sum. In addition, examining the standard deviation, it can be noticed that is less than one in all the cases where the mean is the greatest. The same is also observed for the most of the cases where the standard deviation has the smallest value. Therefore, it is safe to conclude that the sum and mean values can be used to reveal the most important smart device in each case. Comprehensive tables with the evaluation of the devices in each category are presented below (tables 9-11).

Home	Count	Score	Mean	Std dev
Camera (security camera, baby monitor etc.)	**60**	**240**	**4**	**0.91**
Smart lights	59	185	3.14	1.23
Smart thermostat	58	201	3.47	1.09
Virtual Home Assistant (Alexa, google home etc.)	54	165	3.06	1.3
Sensors for room temperature, humidity, air quality etc.	58	220	3.79	0.87
Other (Please specify)	26	79	3.07	1.33

Table 9 - Interval evaluation – Home usage

Environmental	Count	Score	Mean	Std dev
UV sensor	55	215	3.91	1.12
Weather station	**59**	230	3.9	**0.95**
Air quality sensor	56	233	4.16	1.01
Water quality sensor	58	**247**	**4.26**	**0.97**
Other (Please specify)	24	74	3.08	1.29

Table 10 - Interval evaluation – Environmental usage

Smart cities	Count	Score	Mean	Std dev
Parking availability sensor	56	223	3.98	1.06
Traffic camera	**59**	**240**	**4.07**	**0.99**
Charging point	54	198	3.67	**0.98**
Noise sensor	55	178	3.24	1.08
Other (Please specify)	24	63	2.63	1.15

Table 11 - Interval evaluation – Smart cities

However, as Jamieson (2004) stresses out, "treating ordinal scales as interval scales has long been controversial" and some researchers might argue about the validity of the above conclusion.

Considering this controversy, an aggregated examination has been conducted as well to validate the conclusions. In order to create a comprehensive visualisation of the data, clustered bar charts have been created using Excel as shown in figure 2, figure 3 and figure 4 for each category respectively.

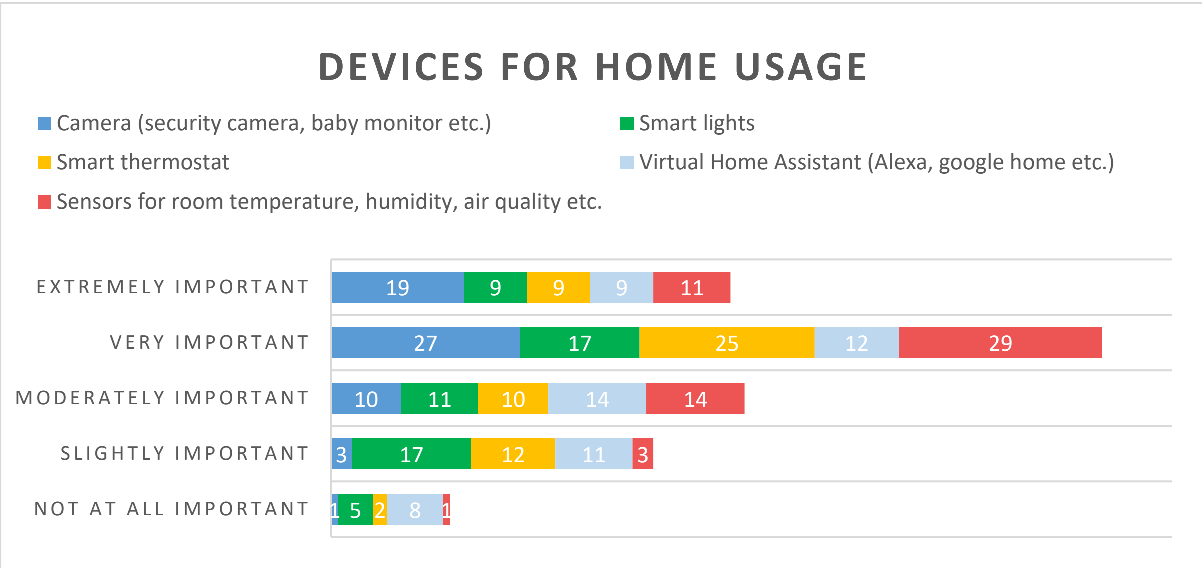


Figure 2 - Aggregated data – Home usage

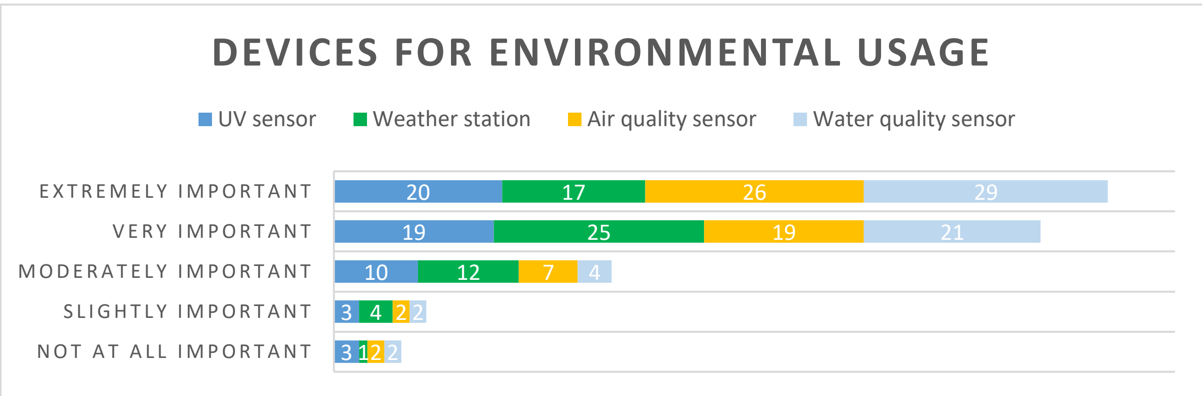


Figure 3 - Aggregated data – Environmental usage

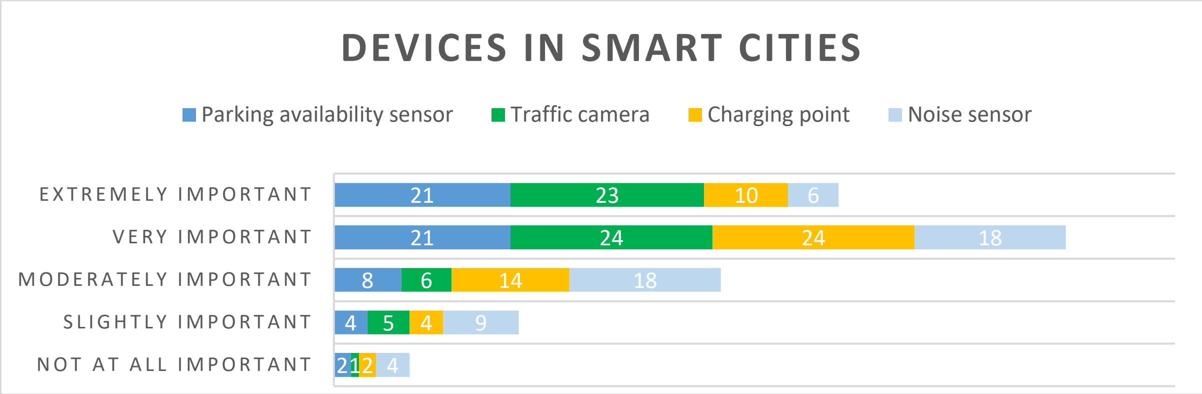


Figure 4 - Aggregated data – Smart cities

The bars represent the number of responses that provided the relevant ranking for each device. As the responses were concentrated in the two highest grades of the scales, it was safely decided to focus on them. Aiming the attention primarily at the 'extremely important' bar, the first conclusions can be made by considering the most important device the one with the greatest frequency of ranking at that level. Following the same approach, we then focus on the 'very important' bar and we examine if the same device is ranked most frequently at that level as well. In the case that the device has the highest frequency of ranking on both level, then it is safe to conclude it is the most important for the users. In the case that there is an ambiguity due to different devices appearing more frequently on different ranking levels, the difference in the number of occurrences at each level is examined. When the difference is greater at the level of 'extremely important', then it derives that the most important smart device is that with the greater number of rankings at the top level (Extremely important).

On the other hand, this approach fails to lead to conclusions when the difference in frequencies is reversed on different ranking levels. In that case the interval approach should be adopted.

Looking at the visualisations and comparing the frequency that each device was ranked as 'extremely important' and 'very important', we conclude that the users consider the camera for the categories of home usage and smart cities, whereas for the devices for environmental use, the participants pointed out that the sensor for measuring water quality is the most important.

3.2.5. Context awareness from the user's perspective

The context awareness section gathered 33 responses in total. 28 were valid while 5 discarded as they contained just numbers or single characters. The qualitative data was analysed by creating themes for each keyword and classifying the responses accordingly. The themes for each search term created as shown in tables 12 - 21:

Air quality:

Theme	Included responses
Air quality sensor/monitor	Devices that are used to simply provide data about the air quality in general without enabling the user to control or change anything.
Pollution sensor	Device that provides information specifically for the air pollution.
Weather station	Device that gathers weather data.
Controller/Actuator	Devices that enable users to act and change the conditions
Alert	Devices that send alerts or notification.
Other	All the rest.

Table 12 - Themes for 'air quality' keyword

Ship:

Theme	Included responses
Geo-locating	Responses that included devices for retrieving the location of a ship. E.g. radar, GPS etc.
Control	Devices enabling users to control part of the ship's equipment.
Monitor	Devices simply for monitoring conditions and functionalities on the ship. E.g. temperature, consumption.
Weather	Devices providing information about the weather.
Other	All the rest.

Table 13 - Themes for 'ship' keyword

Radiation

Theme	Included responses
Radiation sensor/meter	Devices simply for measuring the radiation.
Alert/warning	Devices that send alerts or notifications
UV	Devices measuring UV radiation
Other	All the rest

Table 14 - Themes for 'radiation' keyword

Earthquake

Theme	Included responses
Seismograph/ seismometer	Devices that simply monitor the seismic activity.
Prediction	Devices that can predict earthquakes.
Alert	Devices that send alerts or notifications.
Other	All the rest.

Table 15 - Themes for 'earthquake' keyword

Gamma

Theme	Included responses
Radiation sensor/counter	Devices measuring radiation levels.
Other	All the rest.

Table 16 - Themes for 'gamma' keyword

Weather

Theme	Included responses
Weather station	Weather stations or responses including the word ‘weather’.
Thermometer	Devices simply measuring the temperature.
Thermostat	Devices for controlling the temperature.
Barometer	Barometer.
Forecast	Devices which provide weather forecast.
Wind sensor	Devices for measuring wind-related variables.
Other	All the rest.

Table 17 - Themes for 'weather' keyword

Shark

Theme	Included responses
Sonar/detector	Devices that can detect the presence of a shark
Geo-location	Devices that can provide information about the location of a shark.
Other	All the rest.

Table 18 - Themes for 'shark' keyword

Temperature

Theme	Included responses
Thermometer	Devices for measuring the temperature.
Control	Devices which enable the users to control the temperature.
Weather	Responses for weather.
UV	Responses including the word UV
Other	All the rest.

Table 19 - Themes for 'temperature' keyword

Energy

Theme	Included responses
Monitor	Devices for monitoring and/or measuring energy consumption.
Control	Devices that enable users to control appliances.
Lights	Smart lights or bulbs.
Other	All the rest.

Table 20 - Themes for 'energy' keyword

Aircraft

Theme	Included responses
Geo-location	Responses that included devices for retrieving the location of an aircraft. E.g. radar, GPS etc.
Aircraft status	Devices that provide information about the status in or out of the aircraft. E.g. fuel consumption.
Other	All the rest.

Table 21 - Themes for 'aircraft' keyword

After creating the themes for each search term, the responses were classified accordingly. The occurrences in each theme were counted to conclude at what data the users considered most valuable and should be displayed in each case. The two most important themes with the number of occurrences in each case are shown in tables 22 - 31. The ‘other’ theme has been excluded as it does not provide any useful information. The full report of the responses can be found in appendix 13 and the full report of the thematical analysis in appendix 14.

Air quality:

Theme	Count
Air quality sensor/monitor	15
Controller/Actuator	11

Table 22 - Most important themes - Air quality

Ship

Theme	Count
Geo-locating	14
Control	4
Monitor	4

Table 23 - Most important themes - Ship

Radiation

Theme	Count
Radiation sensor/meter	14
Alert/warning	3

Table 24 - Most important themes - Radiation

Earthquake

Theme	Count
Seismograph/ seismometer	7
Prediction	3

Table 25 - Most important themes - Earthquake

Gamma

Theme	Count
Radiation sensor/counter	13

Table 26 - Most important themes - Gamma

Weather

Theme	Count
Weather station	6
Thermometer	5

Table 27 - Most important themes - Weather

Shark

Theme	Count
Sonar/detector	8
Geo-location	3

Table 28 - Most important themes - Shark

Temperature

Theme	Count
Thermometer	15
Control	8

Table 29 - Most important themes - Temperature

Energy

Theme	Count
Monitor	12
Control	3

Table 30 - Most important themes - Energy

Aircraft

Theme	Count
Geo-location	10
Aircraft status	7

Table 31 - Most important themes – Aircraft

4. Focus groups

After the completion of the online questionnaire and the analysis of the data collected from it, three focus groups conducted. The methodology and the content of the focus groups aimed to gather the users' preferences regarding the interface and the features of an IoT search engine; subsequently, gathering information which would lead to answering the third research question. The focus groups took place at the Andersonian library of the University of Strathclyde, with a total of ten participants. The first two consisted of four participants each and the third consisted of two participants. All the participants were recruited by personal contact as they were affiliated to the researcher through the university. As the research was not funded and the participants were not offered a reward for their participation, there was a limitation in recruiting people who were not related to the researcher and in the total number of participants. The average age of the participants was 27.3, ranging from 22 to 39. At the beginning of the focus groups, all the participants were provided with an information sheet, a consent form and a copy of the pre-search questionnaire (appendices 16, 15 and 167 respectively). The pre-search questionnaire aimed at gathering specific demographic factors (sex and age) of the participants, as well as their familiarity with the Internet of Things search engines. Five of the participants (50%) were female, five (50%) were male and the majority (80%) had not used an IoT search engine before. Instead of keeping notes it was preferred to audio-record the sessions in order not to disrupt the flow of the discussion, to have the ability to review the recordings multiple times to minimise uncaptured information, and finally to keep an archive for future researches. Each focus group lasted approximately 45 minutes.

4.1.Methodology

The focus group was divided in five stages; Introduction, use case scenario, existing IoT search engines evaluation, traditional search engine approach - requirements gathering and final remarks. During each stage, relevant supporting presentation (appendix 21) was displayed on the screen of the room that the focus group conducted. The methodology followed in each stage is described below in sections 4.1.1 – 4.1.5 respectively.

4.1.1. Introduction

This stage was the introductory stage in which the participants were welcomed and informed about the research goals and what was expected from them to do. In addition, the consent forms were distributed, in order to comply with the university's policies, and collected once filled. Following the decided plan, as many users were not familiar with the paradigm of the Internet of Things, a brief description of the concept was given including a cohesive and understandable definition of the IoT. The definition is shown in figure 5 and was chosen as it points out the main drivers of the Internet of Things without including technical details that might be incomprehensible by people without technical background. At that point it is worth mentioning that the participants were recruited regardless their background and experience with smart devices.

A network of devices, equipped with advanced sensors, that are able to collect and exchange data.

Figure 5 - IoT definition in presentation

4.1.2. Search scenario

The participants were provided with the search scenario shown in Figure 6 and were asked to complete a task which included drawing the preferred result page after hypothetically using an IoT search engine to retrieve information from a smart weather station. The purpose of the use of a search scenario was twofold. Primarily, it aimed to encourage the participants to complete the task with greater motivation and secondly to provide the context of the searching purpose to the participants (Tombros, Ruthven and Jose, 2005) in order to determine what is the most useful way of the information to be displayed.

The management of the company you work for, has decided to exploit new technologies and use the information that can be retrieved from smart devices and IoT, to gain competitive advantage. As part of forming the company's strategy, you have been asked to find and document what data can be retrieved from a smart weather station.

To complete the task you use an IoT search engine.

Please draw how you would prefer, the retrieved information to be demonstrated in the result-page.

Figure 6 - Search scenario

Borlund (2000) aptly asserts that the goal of the IIR systems should be to incorporate the information needs of the users as well as to echo the process of seeking and retrieving information. Considering the above, the tests for Information retrieval systems should be carried out by real users so that the assessment of the information retrieved, embody the relevance of each participant individually. On the other hand, the evaluation should be conducted under controlled circumstances so that the results can be compared. Following the latter belief, no other restrictions or instructions were provided regarding the search scenario. However, all the participants were tasked to draw the result page for the same smart device (weather station) to increase the comparability among the responses and to reach to more accurate conclusions. For the choice of the device included in the search scenario, several factors were taken under consideration. Firstly, following the rationale that the smart device should be one of the devices that the users consider as important, the options limited to the five that gathered the highest score in the online survey (water quality sensor, home camera, traffic camera, weather station, air quality sensor). The restrictions and ethical issues that might rise from accessing a camera, led to filtering out the cameras and focusing entirely on the devices used for environmental purposes (water quality sensor, weather station and air quality sensor).The weather station was finally chosen these three devices due to the larger number of devices deployed and being discoverable, the data being publicly accessible as well as the users' familiarity with the type of data that the weather stations generate. The main goal of this stage was to capture the preference of the users regarding the way the information should be presented by the search engine, based totally on their judgement. As the search scenario was at the beginning of the focus group, the participants could include in the drawings, the features of the search engine without being biased by the content of the presentation.

4.1.3. Existing IoT search engines evaluation

As proved from reviewing the literature (section 2), gathering the users' opinion is a common practice for information retrieval systems. Su (1992) identified the "value of the search results" as the most valuable measure for evaluating interactive information retrieval systems. Since, real-world search engines have been implemented already, it was considered crucial to present them to the users and request for their feedback, aiming to form a more comprehensive answer for the third research question. The presentation included using in real-time the existing IoT search engines (Shodan and Thingful) inputting the 'weather station' as a search term. Due the limited number of the results, the 'camera' was used as an additional search term. By using two different search terms, the presentation included a larger and more diverse sample of results as well as more features and connected smart devices. However, the same two search terms were used in all the focus in pursuit of comparable responses. The steps of presenting Shodan and Thingful included:

- Visiting the first page and explaining the features of each search engine
- Typing 'weather station' in the search box
- Reviewing the result page by breaking down the sections of the presented result and explaining the features when needed
- Asking for feedback about what the liked and/or disliked, the usefulness of the results and recommendations.

Acknowledging that the web pages might be offline or inaccessible sometimes, screenshots from the current IoT search engines have been included in the presentation slides as a contingency plan (appendix 21). However, they were not used as both Shodan and Thingful could be reached and browsed.

4.1.4. Traditional search engine approach – requirements gathering

The IoT search engines are still at a primitive stage of development and acceptance. As it was shown from the online survey only a small percentage of the participants knew the existence and an even smaller percentage has ever used one. Hence, simply presenting the existing solutions to the participants of the focus and request feedback about what features they would like to see integrated, was not sufficient. Nevertheless, as the use of the internet expands with a rapidly increasing number of tasks being completed online, the users are very familiar with the traditional search engines. In addition, traditional search engines have developed and incorporated a plethora of different features in order to support the retrieval of the diverse content that can be found online. As a result, in order to fathom which features are preferred by the users (research question 3), the topic was examined using features that already exist in the traditional search engines. Moreover, the fact that most of the features would be recognizable from the participants when presented, acted as another driver to decide to follow that approach. Therefore, the presentation of this stage of the focus group, included images of specific features along with verbal description by the presenter. Each slide of the presentation contained one snapshot for each feature in different search engine (appendix 21). This method was adopted to provide a holistic idea of where each feature might be found and how might be used, as well as to avoid any bias to the participants. The order of presenting the features was chosen randomly. After the presentation, the participants were provided with a list containing all the features presented and asked to rank them based on how important they considered each feature. However, as the list was on paper and only the terminology was written, a slide containing all the snapshots remained on the screen during the ranking task, to provide a visual reminder of what each feature represents. Due to the time limitation, it was impossible to present and collect responses about all the existing features. Even in the case that there were not any time constraints, conducting a long lasting focus group would lead to the fatigue of the participants.

Wilson (2013, pg140) divides the features in four categories. The first category, called input features, includes the components of the search engine interface, which enables the users to indicate their information needs. The following category is the control features; the components which provide the capability to adjust the input. The next set of features, identified as informational, consists of those that display information or metadata about the results. Finally, the features of the search engine which collect and store the user's interactions and other personal details, are classified as personalisable. In this dissertation, the features chosen to be presented and ranked are those that appear to have created a major impact in the way people interact with the search engines. In addition, all four categories were covered.

Arguably, the most valuable component in the input category is the search box. In fact, due to its many advantages, it is considered to be so important for the search task that, as Wilson (2013, pg148) concludes, "searchers can feel at a loss when they do not have a small white text field to spill their search terms into". As a result, it would have no value to include it in the questionnaire and it was considered a requisite feature for any IR system. However, as the smart devices generate diverse data which could be used in various ways and tasks, it was important to include a feature with the capability to assist the users in their searching task. Hence, the facetted search feature, was chosen which provides an enhanced experience when a single query cannot cover an open-ended search task (Wilson, 2013, pg150).

Following to the control category, the filtering and the sorting features have been selected to be included in the presentation and in the ranking list. Filtering and sorting were considered important as they are features widely used and can be found in many different search engines.

From the category of the informational features, only the 'text snippets' features was selected which proved extremely important by reviewing the literature. Research has proved that additional information with each result significantly increases searcher's performance (Dumais, Cutrell and Chen, 2001). Later studies, using eye-tracking technology, have also suggested that the length of the text snippet affects searcher's efficiency in finding the information he was looking for (Laura A. Granka,Joachims and Gay, 2004) and that users spend substantial amount of time in reading the snippets (Laura Granka,Feusner and Lorigo, 2008).

The decision of the personalisation feature in the personalisable category was straight forward as it is the only feature in this category. However, the importance of personalisation is not insignificant since research showed that using a personal profile improves the ranking efficiency (Teevan,Dumais and Horvitz, 2005).

Finally it was decided to include the feature of aggregated search which "aims to facilitate the access to the increasingly diverse content available" (Lalmas, 2011) and has completely changed the way the users look at the information displayed at the result page (Hotchkiss, 2007).

4.1.5. Conclusion and final remarks

At the end of each focus group there was some time allocated for the participants to provide their recommendation, comments or any opinion that they would like to express.

4.2. Results and analysis

4.2.1. Search scenario

It was noticed that, at the beginning, the participants did not know where to start from and asked about more details. The most common questions were if the task was location-specific, if the data should be retrieved from a specific weather station or if they should focus on a particular characteristic of the weather such as the temperature. Another interesting remark was that the participants asked if the data sources were private or public. They were encouraged to include any information they considered useful and it was highlighted that there were no restrictions neither in the content nor in the way the information should be displayed. It is worth pointing out that the participants referred to *Google* multiple times as example in their questions. One participant clearly stated that "The problem is, I am driven by Google". However, to avoid further bias, the presenter avoided referring to *Google* or any other specific search engine. It was highlighted again that were no restrictions, hence if a user considered useful one existing search engine, he could draw it as well.

The participants were given pens and markers of different colours so they could express their thoughts vividly and as a means of promoting their creativity. Additionally, the colourful drawings made easier the identification of each feature and the understanding of the context by the researcher.

After the completion of the focus group the drawings were analysed by breaking down the features that appeared in each drawing and identifying the most important by measuring the frequency of appearances.

It derives that the most important feature it the automatic detection of users' location, which appeared in 6 of the drawings (60%). The participants have either included 'Glasgow' (where the focus groups conducted) in their designs or the more generic 'My location' as in the examples shown in figures 7 and 8. The following most important feature appears to be the capability of retrieving the information needed without following any other link. 5 of the drawings (50%) included representations of weather data, such as temperature and humidity, and weather forecast appearing directly on the result-page. The same percentage (50%) of the participants have demonstrated the feature of the aggregated search in their drawings. For example one participant has included results for proposed activities based on the weather conditions while another participant has included information about "Travel disruptions/warnings". The facets were also identified as a feature in 4 (40%) drawings, giving the ability to the user to choose among the categories of 'location', 'devices' or 'data'. One user designed the preferable result-page solely categorised by "What data can be collected", "What sensors are used to collect the data" and "What format is the data in, and how can it be sent/transmitted to other devices". In 3 (30%) of the drawings, the participants have depicted the capability of having related or suggested searches in the results. The filters were drawn as a feature by the same percentage (30%) of the participants. 2 (20%) of the participants have indicated that the feature of personalisation would be useful by drawing a 'login' option. Finally, 1 (10%) participant has included the option to sort the results while another expressed that a web IoT search engine would not be useful by clearly stating "I'd like an API!". Hence, it is concluded that for the users the most important feature is to retrieve the relevant data just by using the IoT search engine without further action to be needed. Additionally, it was very significant for the users to retrieve the data not only directly from the source, but analysed in a comprehensible manner as well. For example, the users' drawings indicated, that simply retrieving the data that a weather station transmits is not enough. They have expressed the need to see the forecast and the possible effects that derive from the data analysis (i.e. activities to do).

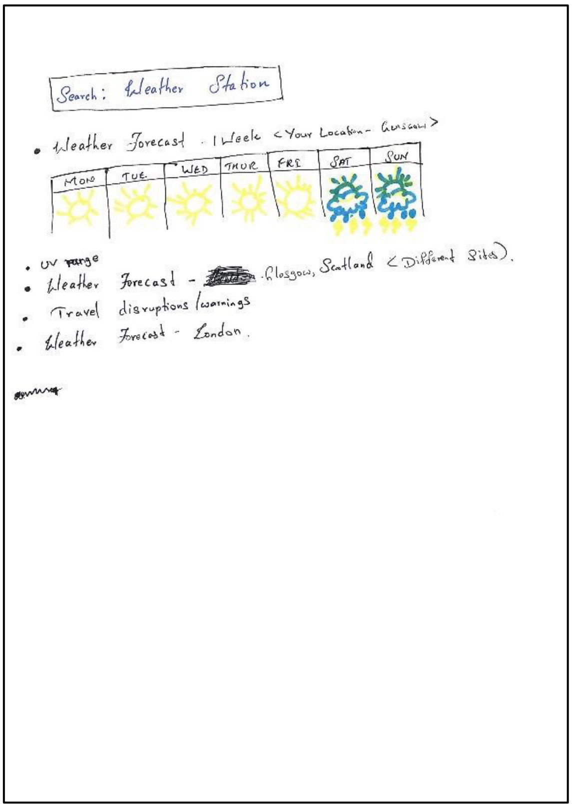


Figure 7 - Focus group drawing (Example 1)

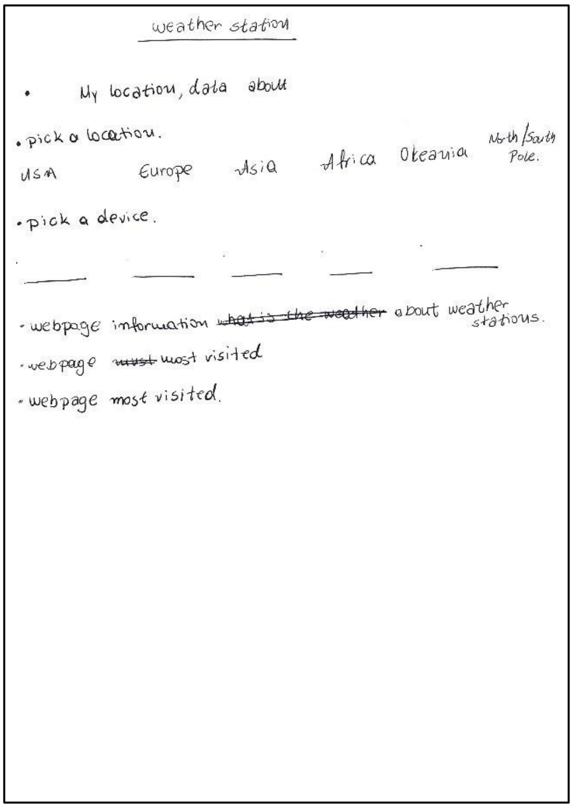


Figure 8 - Focus group drawing (Example 2)

4.2.2. Existing IoT search engines evaluation

After the presentation of each existing IoT search engine (Shodan and Thingful) the participants were asked to comment on the usefulness of the search engine and express their opinions on what they liked or disliked. As previously explained, the sessions were audio-recorded. Hence, for the qualitative analysis the recordings were transcribed in the form of brief notes on the opinions and comments expressed (See appendix 19). Both verbal and non-verbal utterances have been noted for capturing all the information of the recordings. Finally, the punctuation was carefully used were appropriate in order not to change the meaning (Braun and Clarke, 2006). From the thematic analysis of the transcribed data, three major themes emerged: *technical expertise needed, nice but not useful, improvement*. The themes are represented in more detail below.

Technical expertise needed:

Most of the comments classified in this theme referred to Shodan's result page as 6 of the 10 participants have mentioned that it requires high level of technical knowledge which they did not have. For example one participant precisely mentioned that “the technical level of the information you’re getting out is quite high”. It is worth pointing out that some participants were staggered by the result, with one participant just laughing when the result page showed up and stating “Not for us, the common persons” while another, obviously confused asked “How exactly are you using that?”. On the contrary only 1 participant expressed the same opinion for Thingful.

Nice but not useful

This theme mostly includes the comments made for Thingful. Generally, the participants indicated that Thingful’s interface was pleasing (5 participants) and better than Shodan’s (3 participants). For instance 1 participants stated “It’s visually pleasing” while another “This is better”. However, the participants were puzzled about how to use the information from this “cool” interface. On that respect, 2 participants mentioned that “Apart from just getting an overview on what there is out there, I can’t find something useful” and “Nice user interface. But it’s more playing around than getting something useful out of that”. Contrary to the latter, only the comment of one participant can be classified in this theme, who preferred Shodan’s interface as the results were presented in a clear way.

Improvement

From the recorded feedback it is clear that there is the need of improvement of the IoT search engines in order to be considered useful by the general population and be greater adopted. The participants of the focus group in this research appeared to have a good understanding of what additions needed and 5 of the participants provided the relevant recommendations. The recommendations varied from very simple and implementable such as “Filter for public or private. Accessible or not accessible” to more generic as “some analysis. Average temperature, average rain”. Finally, an interesting fact is that 3 participants have identified that the maturity of the IoT and IoT search engines has not reached a satisfactory level yet and the current options are limited.

It appears that the current IoT search engines do not fulfil the requirements for user-acceptance as they lack of features that simplify the information seeking process. People are used to interactive information retrieval systems which provide a variety of data and in most cases already analysed and in a form of useful information. Additionally, the users expected a plethora of control features to manage the search and the results. In accordance with the conclusions of section 4.2.1, it derives that for the users, the most valuable feature of a search engine is to retrieve the relevant information without further action to be needed. The following most important feature appeared to be the aggregated search as the users again expressed the need to have a diverse data which are related to the device they searched for but not explicitly deriving from that. For example one participant suggested to “see data about a fridge and see what extra energy should be consumed on hot days”.

4.2.3. Traditional search engine approach – requirements gathering

Following the methodology described in section 4.1.4, the ranking forms were collected and analysed. The analysis conducted in 4 stages as follows.

1. The conversion scale 6-1 was used, allocating 6 points for the feature in rank 1 as the most important. The next ranks were allocated 5, 4, 3, 2, 1 points respectively (Table 32).
2. The Kendall’s W was calculated to accept or reject the validity of the results.
3. As Kendall’s W revealed that the responses were relatively random and the safe conclusions could not be reached about the importance of the features, another approach followed, using a 8-1 interval conversion scale (table 33).
4. Finally, an adapted conversion scale (table 36) used as a mitigation to the contradictory results of the latter two.

Firstly the simplest method of rank analysis, which is being followed by two of the major survey sites (Qualtrics and Survey monkey) and the scores are allocated to each ranking position following an inverted ordinal scoring. Since the total number of objects was 6 the rank positions were valued as represented in the table below (table 32).

Rank	1	2	3	4	5	6
Allocated score	6	5	4	3	2	1

Table 32 - Search engine rank-to-score conversion scale (Ordinal, 6-1)

Hence, the features were valued based on a 6-1 scale with 6 as the most important feature and 1 the least important. The next step of the analysis included calculating the weighted score of each feature. The weighted score was calculated by summing the total score that each feature gathered, and divide that score by the number of total points in the scoring scale (1+2+3+4+5+6=21) (Savitri Abeyasekera, 2001).

Feature	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**sum**	**Weighted score (sum/21)**
Text snippets	5	3	5	6	5	3	2	3	4	5	41	1.95
Personalised search	6	4	3	1	4	1	1	2	1	2	25	1.19
Sorting	4	2	1	4	1	4	4	4	2	3	29	1.38
Filters	2	1	2	5	2	5	5	6	3	4	35	1.67
Aggregated search	1	6	6	2	5	2	3	1	6	6	38	1.81
Facets	3	5	4	3	3	6	6	5	5	1	41	1.95

Table 33 - Weighted scores (Ordinal, 6-1 scale)

From the final weighted scores (table 33), it appears that the two most important features for the participants were the ‘Text snippets’ and the ‘Facets’ having an equal score of 1.95. The ‘Aggregated search’ feature appears third in the preference of the users but having a score close to the two first.

The Kendall’s W was calculated in order to assess the agreement among the participants and accept or reject the validity of the results. The W factor ranges from 0 to 1; the more the participants have agree on the ranking, the closer the W is to 1. The closer the W statistic is to 0, the more the responses of the participants differ. Considering the responses as seen in table 32, the result of the calculation of W is 0.124. Consequently, there is uncertainty about the validity of the conclusion.

As Maxwell and Bart (1995) assert, a major problem with ranking is that it “provides no information about the spacing between the ranks”. Hence, in the second stage of the analysis, it was followed the approach of scoring the ranks in intervals. The conversion scale that was used to allocate values to the different ranks, was the one proposed by Abeyasekera, Lawson-McDowall and Wilson (1999) for 6 ranks. (table 34)

Rank	1	2	3	4	5	6
Allocated score	8	5	3	2	1	1

Table 34 - Search engine rank-to-score conversion scale (Interval, 8-1) (S. Abeyasekera,Lawson-McDowall and Wilson, 1999)

As it can be observed, the main difference from the conversion scale in method 1, is that the differences between the scores in each rank, are greater in the higher ranking positions. Applying the latter conversion to the data and following the same calculation procedure, the features appeared in a different order of importance (table 35).

Feature	**Sum**	**Weighted score (sum/20)**
Aggregated search	43	2.15
Facets	41	2.05
Text snippets	38	1.9
Filters	32	1.6
Personalised search	22	1.1
Sorting	21	1.05

Table 35 - Weighted scores (Interval, 8-1 scale)

Hence, by using an interval conversion scale which allocates more weight to the first ranks, it appears that the participants considered the ‘aggregated search’ as the most important, while ‘facets’ and ‘text snippets’ ranked in the second and third position respectively. Finally, the ‘filters’ feature remained in the fourth position of users’ preferences while ‘personalised search’ and ‘sorting’ features swapped in the last positions. It is worth pointing out, that the results of this method are in accordance with the conclusions made by analysing the users’ drawings in section 4.2.1.

The results from the qualitative analysis using two different methods do not allow to accurately conclude about the importance of the search engine features. The controversy of the results proves that the responses of the participants do not follow any trend and vastly differ. To mitigate the controversy of the results, a third conversion scale was used with an ad hoc adaptation. In the adapted conversion scale, 21 points in total were allocated per column (participant) as in the first method. However, the interval approach was followed in the distribution of the 21 points, with the top ranks having greater weight (table 36).

Rank	1	2	3	4	5	6
Allocated score	7	5	4	3	1	1

Table 36 - Search engine rank-to-score conversion scale (Interval, 7-1 scale)

The first conversion scale used the smallest difference (using integers) between the ranks. As a result the final ranking was highly dependent on the frequency of responses in high or middle ranking positions. On the other hand, the results using the second scale were vastly affected by the times a feature was ranked in the first position as 40% of the points (8 out of 20) were allocated solely in the top rank. Taking under consideration the small number of the participants (10) it was considered essential to create a mitigation conversion scale and compare the results from the three methods. The same procedure followed to calculate the weighted scores of the rankings as shown in table 37.

Feature	**Sum**	**Weighted score (sum/21)**
Facets	43	2.05
Text snippets	41	1.95
Aggregated search	40	1.90
Filters	33	1.57
Sorting	27	1.29
Personalised search	24	1.14

Table 37 - Weighted scores (Interval, 7-1 scale)

The 'facets' appears to be the most important feature based on the results of the third method, with a weighted score of 2.05. The following most important features appear to be the 'text snippets' and the 'aggregated search' with a weighted score of 1.95 and 1.90 respectively.

Even though three different methods used to discover which features the users consider as the most important, all of the methods led to different results. Consequently, it cannot accurately derive a ranked list of the features that the users would prefer. However, it is worth mentioning that the top three features can be identified as the 'facets' aggregated search' and 'test snippets' which remained in the top three positions regardless the method used. In addition, it can be concluded that the 'filters' feature is considered fourth in preference while the last two are the 'personalised search' and the 'sorting'.

5. Conclusion and recommendations

It should be mentioned that some limitations were identified during the completion of this research. The unavailability of public data by smart devices can be identified as the most important. The lack of publicly available data using Shodan and Thingful, did not allow deeper exploration of their usability, neither thorough assessment by the users. Additionally, the research was conducted in part fulfilment of requirements for the degree of MSc Information Management, hence it was not funded. As a result, there was a limitation in the number of recruited participants as there was not any reward for the participation. Finally, it should be prompted that the research should be completed in the period of 12 weeks, which was the time frame for the master's programme dissertation. Consequently, the online questionnaire could not remain active for longer to gather more responses and there was no time for more focus groups to be conducted. Finally, the limitations that rise from the fact that the thematic analysis is sensitive to subjectivity, are recognised.

In this research the first steps for a user-centred design have been made for the IoT search engines. Further research is recommended to capture the preferences of a larger audience as well as to focus on specific target groups where necessary. For example, as there is a lot of research of how the IoT could be used to assist older people, conducting a similar study focused on the preferences of older users could benefit the adoption of IoT by them. Finally, as the businesses represent 57% of overall IoT spending (Van der Meulen, 2017), it is recommended that other studies should focus on the preferences of the users in a working environment.

This research has been conducted with the purpose to cover the gap that has been identified in the literature (section 2) and create a baseline for further studies. The research questions (section 1) that defined and answered by this study, aimed to acquire deeper knowledge of the user behaviour and needs.

The first objective of the research was to explore the user understanding and adoption rate of the IoT paradigm. To achieve this objective, an online questionnaire was launched and distributed via the social media. The analysis of responses indicated that 90% of the respondents had a relatively solid understanding of what internet of things is. Additionally, it was revealed that the adoption rate of IoT is quite high as 60% of the respondents stated that they have used a smart device. On the contrary, the rate of the respondents who knew about the existence of an IoT search engine such as Shodan or Thingful, dropped to 30.4%. Hence, it becomes obvious that the IoT search engines need to be improved to reach the same adoption rate with the IoT paradigm in general.

The second objective was to understand what information is the most valuable for the users (research question 2) in order to retrieve the most relevant data from an IoT search engine. A single query cannot accurately define the information needs of a user, hence, knowing the importance of each data set will assist in the implementation of an IoT search engine which can provide more relevant results to the user. Moreover, the appropriate indexing of the smart devices is considered crucial to define the context of the information that they transmit. As a result, context awareness from the user perspective has been researched as well. The second part of the questionnaire (sections 3.1.4 - 3.1.5 and 3.2.4 - 3.2.5), aimed to complete the aforementioned objective. The users where asked to evaluate the importance of data retrieved from different smart devices in three categories (home, environment and smart city) using Likert-type scales. The results revealed that the users consider as the most important the data from smart home cameras and traffic cameras. The data from water quality sensors and room conditions sensors have also been identified as very important from the users. Thematic analysis conducted for the context awareness, from which it derives that most users expect to retrieve data from device which monitor the physical environment.

Finally, aiming to identify user-preferable features of an IoT search engine, three focus groups were conducted. The results presented the real problem of the current IoT search engines; they are not user-friendly and a high level of technical knowledge is required to use them. In pursue of identifying the features that should be included in a user-centred design (Abras, Maloney-Krichmar and Preece, 2004), the users requested to draw how the information from a weather station should be displayed and then they were requested to rank six of the most important features of the traditional search engines. The analysis showed that the relevant information should be immediately been displayed after the query is submitted, without requiring any further action. In addition, the users indicated that the most valuable features were the 'aggregated search', the 'snippets' and the 'facets' , but not in a definitive order.

Summing up, in this research a user-centred approach has been followed for the IoT search engines which proved that despite the fact that the users have adopted the IoT paradigm, the respective information retrieval systems do not sufficiently cover the users' needs. For a greater adoption, the IoT search engines should incorporate more of the features that the users are familiar with. Based on the findings of this research, the features that should be in top priority in the implementation of an IoT search engine are the:

No-click: Automatically and immediately display the relevant information on the result page.

Aggregated search: Provide information from multiple and diverse sources of data.

Facets: Giving the ability to the user to choose from different categories.

Text snippets: Display information from within the source, below the source link.

Furthermore, focusing on the importance of the diversity of the information displayed to the users, the research has also revealed that relevant information is not considered only the information retrieved from a well-indexed smart device. The participants in the focus have designated that it is also important to retrieve information from other sources not directly related to the query but linked to the requested information; for example retrieving energy data that could be related to the weather conditions (see appendix 20, focus group 2, participant 1). Hence, the context of each devices along with the way it is associated with other devices, should be taken under consideration when designing an IoT search engine.

Bibliography

Abeyasekera, S. (2001) 'Analysis approaches in participatory work involving ranks or scores'. *DFID Theme Paper (revised). UK: Statistical Services Centre, University of Reading*.

Abeyasekera, S., Lawson-McDowall, J. and Wilson, I. (1999) 'CONVERTING RANKS TO SCORES FOR AN AD-HOC ASSESSMENT OF METHODS OF COMMUNICATION AVAILABLE TO FARMERS'. *Methodological Framework Integrating Qualitative and Quantitative Approaches for Socio-Economic Survey Work, Theme Paper,* 4 6.

Abras, C., Maloney-Krichmar, D. and Preece, J. (2004) 'User-centered design'. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications,* 37 (4), pp. 445-456.

Al-Maskari, A., Sanderson, M. and Clough, P. (2007) Published. 'The relationship between IR effectiveness measures and user satisfaction'. 2007 2007. ACM, pp.773-774.

Ashton, K. (2011) 'That 'internet of things' thing'. *RFiD Journal,* 22 (7), pp.

Bandyopadhyay, D. and Sen, J. (2011) 'Internet of Things: Applications and Challenges in Technology and Standardization'. *Wireless Pers Commun,* 58 (1), pp. 49-69.

Belkin, N.J. *et al.* (1995) 'Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems'. *Expert Systems With Applications,* 9 (3), pp. 379-395.

Belkin, N.J., Marchetti, P.G. and Cool, C. (1993) 'BRAQUE: Design of an interface to support user interaction in information retrieval'. *Information Processing and Management,* 29 (3), pp. 325-344.

Bodenheim, R. *et al.* (2014) 'Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices'. *International Journal of Critical Infrastructure Protection,* 7 (2), pp. 114-123.

Boone, H.N.J. and Boone, D.A. (2012) 'Analyzing Likert Data'. *Journal of Extension,* 50.

Borlund, P. (2000) 'Experimental components for the evaluation of interactive information retrieval systems'. *Journal of Documentation,* 56 (1), pp. 71-90.

Borlund, P. (2009) 'User-centered evaluation of information retrieval systems'. *Information Retrieval: searching in the 21st Century,* 21-37.

Botta, A. *et al.* (2016) 'Integration of Cloud computing and Internet of Things: A survey'. *Future Generation Computer Systems,* 56 684-700.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology'. *Qualitative Research in Psychology,* 3 (2), pp. 77-101.

Cavalcante, E. *et al.* (2016) 'On the interplay of Internet of Things and Cloud Computing: A systematic mapping study'. *Computer Communications,* 89-90 17-33.

Couper, M.P., Traugott, M.W. and Lamias, M.J. (2001) 'Web Survey Design and Administration'. *Public Opinion Quarterly,* 65 (2), pp. 230-253.

Ding, Z., Chen, Z. and Yang, Q. (2014) 'IoT- SVKSearch: a real-time multimodal search engine mechanism for the internet of things'. *International Journal of Communication Systems,* 27 (6), pp. 871-897.

Dubbeldeman, R. and Ward, S. (2015) 'Smart Cities, How rapid advances in technology are reshaping our economy and society'. Available at: https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/public-sector/deloitte-nl-ps-smart-cities-report.pdf.

Dumais, S. *et al.* (2016) 'Stuff I've Seen: A System for Personal Information Retrieval and Re-Use'. *ACM SIGIR Forum,* 49 (2), pp. 28-35.

Dumais, S., Cutrell, E. and Chen, H. (2001) Published. 'Optimizing search by showing results in context'. 2001 2001. ACM, pp.277-284.

Ellis, D. (1989) 'A Behavioural Approach to Information Retrieval System Design'. *Journal of Documentation,* 45 (3), pp. 171-212.

Genge, B. and Enăchescu, C. (2016) 'ShoVAT: Shodan-based vulnerability assessment tool for Internet-facing services'. *Security and Communication Networks,* 9 (15), pp. 2696-2714.

Granka, L., Feusner, M. and Lorigo, L. (2008) 'Eye Monitoring in Online Search'. In: Hammoud, R.I. (ed.) *Passive Eye Monitoring: Algorithms, Applications and Experiments.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 347-372.

Granka, L.A., Joachims, T. and Gay, G. (2004) Published. 'Eye-tracking analysis of user behavior in WWW search'. 2004 2004. ACM, pp.478-479.

Göb, R., McCollin, C. and Ramalhoto, M. (2007) 'Ordinal Methodology in the Analysis of Likert Scales'. *Qual Quant,* 41 (5), pp. 601-626.

Hart, M.C. (1996) 'Improving the discrimination of SERVQUAL by using magnitude scaling'. *Kanji,G.K.(1996) : TQM In Action! (Chapman and Hall).*

Hopfgartner, F. *et al.* (2008) Published. 'Search trails using user feedback to improve video search'. 2008. ACM, pp.339-348.

Hotchkiss, G. (2007) *Eye Tracking On Universal And Personalized Search.* Available at: http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12233.

Hu, P.J.-H., Ma, P.-C. and Chau, P.Y.K. (1999) 'Evaluation of user interface designs for information retrieval systems: a computer- based experiment'. *Decision Support Systems,* 27 (1), pp. 125-143.

Jamieson, S. (2004) 'Likert scales: how to (ab) use them'. *Medical Education,* 38 (12), pp. 1217-1218.

Lacey, A. and Luff, D. (2007) *Qualitative data analysis* [Press release].

Lalmas, M. (2011) 'Aggregated Search'. In: Melucci, M. and Baeza-Yates, R. (eds.) *Advanced Topics in Information Retrieval.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 109-123.

Lee, P. and Talbot, E. (2016) 'The Internet of Things: most entertaining'. Available at: https://www.deloitte.co.uk/mobileuk/assets/pdf/Deloitte-Mobile-Consumer-2016-There-is-no-place-like-phone.pdf.

Li Da, X., Wu, H. and Shancang, L. (2014) 'Internet of Things in Industries: A Survey'. *Industrial Informatics, IEEE Transactions on,* 10 (4), pp. 2233-2243.

Lodge, M. (1981) 'Magnitude scaling: quantitative measurement of opinions'.

Lunardi, W.T. *et al.* (2015) Published. 'Context-based search engine for industrial IoT: Discovery, search, selection, and usage of devices'. 2015 2015. IEEE, pp.1-8.

Maxwell, S. and Bart, C. (1995) 'Beyond ranking: exploring relative preferences in P/RRA'. *PLA Notes,* 22 28-34.

National Intelligence Council (2008) Published. 'Six Technologies with Potential Impacts on US Interests out to 2025'. *Disruptive Civil Technologies,* 2008.

Ostermaier, B. *et al.* (2010) Published. 'A real-time search engine for the Web of Things'. *2010 Internet of Things (IOT),* Nov. 29 2010-Dec. 1 2010 2010. pp.1-8.

Perera, C. *et al.* (2014) 'Context Aware Computing for The Internet of Things: A Survey'. *IEEE Communications Surveys & Tutorials,* 16 (1), pp. 414-454.

Petrelli, D. *et al.* (2004) 'Observing Users - Designing clarity a case study on the user-centred design of a cross-language information retrieval system'. 923-934.

Rivera, J. and Van der Meulen, R. (2015) *Gartner's 2015 Hype Cycle for Emerging Technologies Identifies the Computing Innovations That Organizations Should Monitor* [Press release].

Shemshadi, A., Sheng, Q.Z. and Qin, Y. (2016) Published. 'Thingseek: a crawler and search engine for the internet of things'. 2016 2016. ACM, pp.1149-1152.

Shemshadi, A., Sheng, Q.Z. and Qin, Y. (2017) '*The Anatomy of An Intent Based Search and Crawler Engine for the Web of Things'*.

Shemshadi, A. *et al.* (2016) 'Searching for the Internet of Things on the Web: Where It Is and What It Looks Like'. *arXiv preprint arXiv:1607.06884*.

Shodan. (2017) Available at: http://www.shodan.io.

Sikkel, D., Steenbergen, R. and Gras, S. (2014) 'Clicking vs. Dragging'. *Public Opinion Quarterly,* 78 (1), pp. 177-190.

Su, L.T. (1992) 'Evaluation measures for interactive information retrieval'. *Information Processing and Management,* 28 (4), pp. 503-516.

Tan, C. *et al.* (2010) 'Microsearch: A search engine for embedded devices used in pervasive computing'. *ACM Transactions on Embedded Computing Systems (TECS),* 9 (4), pp. 1-29.

Teevan, J., Dumais, S.T. and Horvitz, E. (2005) Published. 'Personalizing search via automated analysis of interests and activities'. 2005 2005. ACM, pp.449-456.

Tombros, A., Ruthven, I. and Jose, J.M. (2005) 'How users assess Web pages for information seeking'. *Journal of the American Society for Information Science and Technology,* 56 (4), pp. 327-344.

Uckelmann, D., Harrison, M. and Michahelles, F. (2011) 'An architectural approach towards the future internet of things'. *Architecting the internet of things.* Springer, pp. 1-24.

Van der Meulen, R. (2017) *Gartner Says 8.4 Billion Connected Things Will Be in Use in 2017, Up 31 Percent From 2016* [Press release].

Wang, H., Tan, C.C. and Li, Q. (2008) '*Snoogle: A search engine for the physical world*'.

Wilson, M. (2013) 'Interfaces for Information retrieval'. In: Kelly, D. and Ruthven, I. (eds.) *Interactive Information Seeking behaviour and retrieval.* London: Facet Publishing, pp. 139-170.

Yap, K.-K., Srinivasan, V. and Motani, M. (2008) 'MAX: Wide area human- centric search of the physical world'. *ACM Transactions on Sensor Networks (TOSN),* 4 (4), pp. 1-34.

Appendices

Appendix 1 – Submission for ethics approval

Application ID: 606

Title of research:
User-preferable features in an Internet of Things search engine.

Summary of research (short overview of the background and aims of this study):
This research is about a relatively new paradigm, the Internet of Things (IoT). As the technology progresses and the devices connected to the internet exponentially increases and affects the way people live, new tools are needed to harness the data these devices collect. In this respect, search engines specifically designed for serving the IoT paradigm have been created. However, they are still not widely adopted by the majority of the internet users. Even though extensive research exists about the engineering and the frameworks that IoT search engines operate in, no literature can be found about the way the users would like to interact with such a search engine. Consequently, this project aims to examine what are the features that users would like, cover a part of the gap in literature about the usability of these search engines and provide a basis for further user-centred research.

How will participants be recruited?
The participants of the focus groups will be recruited online, after completing the survey. In the last section they can voluntarily insert their email to be invited at the focus group.

How will consent be demonstrated? Either upload or include here a copy of the consent form/instructions issued to participants.
PDF File: View document
In the online survey, the consent will be retrieved at the introduction (Please see the link of the survey).
The consent form of the focus group is attached.

What will the participants be told about the proposed research study? Either upload or include a copy of the briefing notes issued to participants.
PDF File: View document
See attached.

What will participants be expected to do? Either upload or include a copy of the instructions issued to participants along with a copy of or link to the survey, interview script or task description you intend to carry out. Please also confirm (where appropriate) that your supervisor has seen and approved both your planned study and this associated ethics application.
PDF File: None.
PDF File: None.
Online survey: https://strathsci.qualtrics.com/jfe/form/SV_6EupMWh0xeJIZYF
Focus group: They will be shown some existing features of different search engines and asked to evaluate them and comment on their usability. In addition, some examples about information retrieval results will be presented and asked to evaluate them and comment. Finally, they would be asked about recommendations for possible improvements.

What data will be collected and how will it be captured and stored? In particular indicate how adherence to the Data Protection Act will be guaranteed and how participant confidentiality will be handled.
In the online survey only the email and name will be collected in order to invite the participant in the focus group. The data will be stored in the departmental server in which the Qualtrics tool stores information. No copies to personal hard drives or cloud services will be made.
The data collected in the focus groups, will be the data included in the consent form. The consent forms, will be given to the department and will be handled according to the procedures and regulations of the University of Strathclyde. No copies will be kept by the researcher.

How will the data be processed? (e.g. analysed, reported, visualised, integrated with other data, etc.)
The data gathered online will be stored and processed in the university server, in which qualtrics stores the data.
The data gathered in the focus groups will also be securely stored in the university's server where the researcher (Dimitrios Markogiannis) and the supervisor (Dr Martin Halvey) have access.

How and when will data be disposed of?
At the end of the research the student will delete any data from his personal H: hard drive. No other hard drive or cloud service will be used.
It is at the university's discretion of how the data stored in the University's archive, are going to be disposed.

Appendix 2 – Ethics approval received

8/11/2017 Mail – dimitrios.markogiannis.2016@uni.strath.ac.uk

Ethics application has been approved

www-data <www-data@cis.strath.ac.uk>

Thu 29/06/2017 19:28

To:Dimitrios Markogiannis <dimitrios.markogiannis.2016@uni.strath.ac.uk>;

Hello,

Your ethics application "User-preferable features in an Internet of Things search engine." (ID: 606) has been approved.

URL: https://local.cis.strath.ac.uk/wp/extras/ethics/index.php?view=606

Ethics Approval System.

Appendix 3 – Online questionnaire

IoT Search Engine

Survey Flow

Standard: Aim of research / consent (1 Question)
Block: Demographics 1 (1 Question)
Standard: Demographics 2 (1 Question)
Standard: General background (1 Question)
Standard: IoT (1 Question)
Standard: Search engine (5 Questions)
Standard: Smart devices (2 Questions)
Standard: Devices per catergory 1 (1 Question)
Standard: Devices per category 2 (1 Question)
Standard: Devices per category 3 (1 Question)
Standard: Context sense (1 Question)
Standard: Focus group participation (1 Question)

Aim of research / consent

Q1.1 Thank you for taking part in my survey about the Internet of Things search engines.

The following questions aim to research the familiarity of the users with the concept of Internet of Things (IoT), how it is perceived and to discover how the search engines that are specifically designed for supporting the IoT paradigm, can be improved and be used by a broader range of users.

This survey complies with the University of Strathclude data privacy policy and all the data collected will be stored in the file-store service 'StrahCloud'. Providing your contact details at the end of this survey is voluntary and will be used only to contact you to participate in a follow-up focus group which will provide richer content to the research. You retain your right to have your data removed if you wish. Please note that this is possible only if you provide contact details, otherwise it is not possible to identify your responses.

The survey will take approximately 5-10 minutes. You can quit this survey any time you wish by closing the tab of the browser or the browser itself. Please note in case of any questions, queries or doubts, you can always directly contact the researcher of the sypervisor using the details below.

Creator: Dimitrios Markogiannis	Email: dimitrios.markogiannis.2016@uni.strath.ac.uk
Project Supervisor: Dr Martin Halvey	Email: martin.halvey@strath.ac.uk

By clicking the '>>' button below you consent to store and analyse your responses for the purposes of this research project.

End of Block

Demographics 1

Q2.1 Gender

- Male (1)
- Female (2)
- Other (3)
- Prefer not to say (4)

End of Block

Demographics 2

Q3.1 Please specify your age.

End of Block

General background

Q4.1 What is your background related to Information Technology/Computer Science?

- Academic (E.g. student, researcher) (1)
- Professional (2)
- No background related to IT/Computer Science (3)

End of Block

IoT

Q5.1 Please explain in your own words what do you understand to be the meaning of "Internet of things".

End of Block

Search engine

Q6.1 Similarly to the search engines that are used to find content in the internet (e.g. Google, Yahoo!, Bing etc), there are Internt of Things search engines (e.g. Shodan, Thingful) that are used to find smart devices connected to the internet and provide relevant data to the user. Did you know about the existence of such search engines?

- ○ Yes (1)
- ○ No (2)

Skip To: End of Block If Q6.1 = No (2)

Q6.2 Have you ever used an Internet of Things search engine?

- ○ Yes (1)
- ○ No (2)

Display This Question:

If Have you ever used an Internet of Things search engine? Yes Is Selected

Q6.3 Which was the last IoT search engine that you have used?

Display This Question:

If Have you ever used an Internet of Things search engine? Yes Is Selected

Q6.4 What term(s) did you search for? Please list up to 3 and press 'enter' after each term.

Display This Question:

If Have you ever used an Internet of Things search engine? Yes Is Selected

Q6.5 How satisfied were you with the information retrieved?

- ○ Extremely satisfied (1)
- ○ Somewhat satisfied (2)
- ○ Neither satisfied nor dissatisfied (3)
- ○ Somewhat dissatisfied (4)
- ○ Extremely dissatisfied (5)

End of Block

Smart devices

Q7.1 A smart device is a device that can be remotely controlled and/or monitored when connected to the Internet or a local network. E.g. Smart thermostat, smart lights, IP camera, wearable device etc.Have you ever used such a device?

- ◯ Yes (1)
- ◯ No (2)

Display This Question:

If A smart device is a device that can be remotely controlled and/or monitored when connected to the Internet or a local network. E.g. Smart thermostat, smart lights, IP camera, wearable device etc.Ha... Yes Is Selected

Q7.2 What device(s) was/were that? Please list up to 3 that you have used. (Please separate by pressing 'enter' after each device)

____

____

____

____

End of Block

Devices per category 1

Q8.1 The devices below are designed for home usage. How important do you consider to be able to discover each of the following devices and retrieve the data that it produces?

Camera (Security camera, baby monitor etc) (1)	
Smart lights (2)	
Smart thermostat (3)	
Virtual home assistant (Alexa, google home etc) (4)	
Sensors for room temperature, humidity, air quality etc (5)	
Other (Please specify) (6)	

End of Block

Devices per category 2

Q9.1 The devices below are used for environmental purposes. How important do you consider to be able to discover each of the following devices and retrieve the data that it produces?

UV sensor (1)	
Weather station (temperature, humidity, wind etc) (3)	
Air quality sensor (4)	
Water quality sensor (5)	
Other (Please specify) (6)	

End of Block

Devices per category 3

Q10.1 The devices below are used in smart cities. How important do you consider to be able to discover each of the following devices and retrieve the data that it produces?

Sensor for parking availability (1)	
Traffic camera (3)	
Charging point (4)	
Noise sensor (5)	
Other (Please specify) (6)	

End of Block

Context sense

Q11.1 Below are listed the 10 most popular keywords in IoT search engines. Please fill in up to 3 devices that you would expect to find by searching each keyword in an IoT search engine. (Separate the devices by comma)

Keywords Expected devices

○ air quality (1) ____

○ ship (2) ____

○ radiation (3) ____

○ earthquake (4) ____

○ gamma (5) ____

○ weather (6) ____

○ shark (7) ____

○ temperature (8) ____

○ energy (9) ____

○ aircraft (10) ____

End of Block

Focus group participation

Q12.2 This research also consists of focus groups in which the participants will be asked to evaluate current IoT search engines, comment on the existing and potential features, and provide their recommendations. If you would like to participate in one of the focus groups, which will take place in the campus of the University of Strathclyde, please fill in the form below.

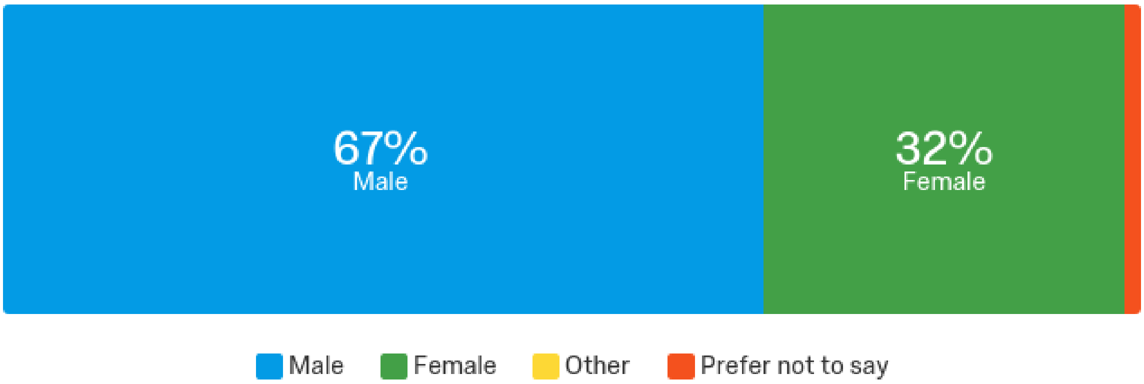
- Name (1) ____

- Email (2) ____

End of Block

Appendix 4 – Online survey: Gender report

Q2.1 - Gender



#	Answer	%	Count
1	Male	66.67%	46
2	Female	31.88%	22
3	Other	0.00%	0
4	Prefer not to say	1.45%	1
	Total	100%	69

Appendix 5 - Online survey: age statistics

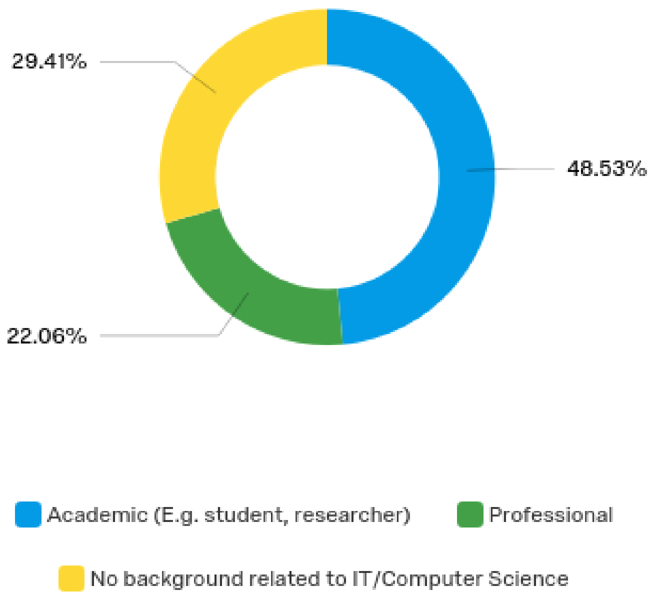
Q3.1 - Please specify your age.

Age	
22	29
22	29
23	29
24	29
25	29
25	29
25	29
25	30
25	30
25	30
25	30
25	31
26	31
26	32
26	32
27	32
27	33
27	33
27	33
27	35
27	35
27	35
27	36
28	37
28	37
28	38
28	39
28	40
28	40
28	40
29	41
29	43
29	52
29	61
29	

group 18-25:	12
group 26-35:	45
group > 35:	12
Mean:	30.65217
Median:	29
Min:	22
Max:	61

Appendix 6 – Online survey: Users' background

Q4.1 - What is your background related to Information Technology/Computer Science?



#	Answer	%	Count
1	Academic (E.g. student, researcher)	48.53%	33
2	Professional	22.06%	15
3	No background related to IT/Computer Science	29.41%	20
	Total	100%	68

Appendix 7 – Online survey: Users' familiarity with IOT

Q5.1 - Please explain in your own words what do you understand to be the meaning of "Internet of things".

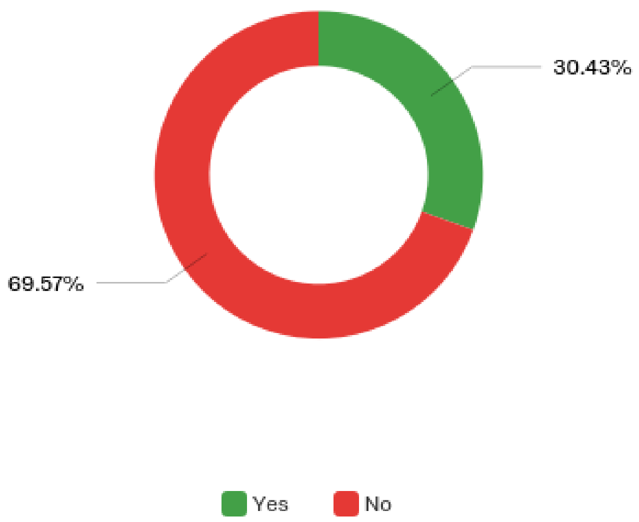
It is a "system" in which all devices are connected to each other and are able to send and receive data
Information
connecting several devices through the internet in order to exchange data
It is the interconnection of mobile devices via the internet, these devices have personal internet protocol (IP) addresses with which they are identified with and it allows them to carry out tasks automatically.
I have no idea
Connect electronic devices to a public and/ or private network.
Connected devices, every day objects communicating with one another and with us through the internet
connection of daily life products like toasters, windows, heating etc
Connecting devices such as vehicles, fixtures and fittings via the internet and enabling us to share and exchange data.
Interconnection between all entities forming a gaint network
Any kind of device can be connected to the internet. Via the internet it can be controlled, adjusted, or analysed.
A network of physical devices, buildings, vehicles, and other items of any sort that are implanted with a type of software, with the purpose of collecting data.
micro devices with the ability to connect on internet
The networking of smart devices
Internet of things, is the trend to give internet connection in devices used everyday, such as cars, buildings,elevetarors, etc.
It is a new concept that allows the users to get information from devices, which are connected to the internet.
Item that operate using internet connection
Ordinary items which are connected to the Internet in order to improve performance or usability such as smart tvs or refrigerators that can order food when it runs out
Using of internet for devices to collect and exchange data
internet of things are digital embedded devices that are smart and capable to communicate over the cloud services.
The connection over a network of devices in order to communicate and exchange data
Something related to internet
Things that you are doing in the internet. Surfing, research, etc. Information for a lot of things that you are interested in.
Ways to connect disparate devices to the internet, either directly or via a network, in order that they can be centrally monitored or controlled for the overall benefit of an environment or the groups and individuals in it.
According to me, things on the internet are all inter netted. All things are netted together in some ways.
All the imformtions there are one the web
everyhting connected to the internet
connected products/objects, machine-to-machine communication, etc.
Embedded systems connected to the internet

It's how appliances such as fridges, light bulbs, boilers etc are connected to web services in order to update their firmware, to be controlled by user, to communicate with other appliances or service providers and so on.
Machine to machine communication, connected devices that offer never before seen information through their interfaces, AI, sensors, etc
little machines collecting data. a server collects data from many resources , forwards them to another server for data analysis, based on the analysis , decisions can be made to maximize the profitability of a product. Another example. every device can be part of internet.
Enabling connectivity in services, industry and infrastructure appliances and products that teaditionally wouldn't be connected to a network. This allows collection and realtime processing of data on operationg conditions and performance that can be used to quickly obtain actionable insights into the way these products operate. This will result in better performance, higher flexibility and better customer experience.
Everything will have an IP and will be accessible via net
Internet of thing is the connection of things necessary or not to the internet
All devices are connected to the network providing and receiving useful information for automation handling or user's convinience.
Connection of things
Things: devices such as washing machine and traffic lights Internet: Currently it refers to the Web that connects computers and mobile devices. Internet of things: Internet that connects the "things" as described above.
every device "speaks" the language of internet
A popular term for the concept of inter- or internet-connected devices, not used for communication or media consumption, used where traditionally a non-connected devices was used. A well known case is using those devices at home 'smart home', a typical iot devices is a Nest thermostat. (An iPad or Smart TV is generally not considered to be an iot device)
devices connected to the Internet
Connecting "things" online in order to remotely control them or/and program them so they automatically "act" upon certain conditions
Deploy of connectivity everywhere to sense different "things" in quasi-real time.
Devices connected to the web that we then control from our phone or other web interface
It refers to internet-connected everyday devices - the Internet connection of almost anything that isn't a traditional computing device (e.g. a refrigerator)
Using of IT and Internet to connect devices for using, collecting, and exchange data between devices
Network/ connection
Its a network integrated by intercommunicated devices that interact with each-other and with the network itself without the intervention of a human user, in principle.
The interconneciton of Cyberphysical systems for an easier and more convenient future
Internet of things means that all devices, which have a network connection possibility, are able to exchange the information with each other through the Internet or internal network.
With the IoT we have the ability to interconnect devices, bigger or smaller, vehicles, buildings, houses and in general every item that can use electronics, sensors and software. This connection can give us a huge amount of information and data that can be processed and exchanged among the participants in order to make our daily life easier and more enjoyable.
Everything that you use and have is connected to the Internet. In plain words, you can see what your children do in the home by using camera or where is your dog through GPS necklace e.g,
Adding intelligence to sense, collect, react, and communicate in a every day objects.
It is a paradigm of interconnected objects

A collection of services and information made available to people with technology devices through a collection of networks
For me, Internet of Things is the inter-connection (network) of objects (computing devices) via internet in order to send and receive data.
How internet affects our daily life.
Network of devices connected to Internet that can communicate with (send data to) each other.
An application model where devices and cyberphysical systems have computational capabilities and are connected to the Internet.
It is a potentially huge network of various big/small devices connected together through some protocol similar to TCP/IP.
How we're all connected, ie smart phones, smart watches, smart tvs! Cant go anywhere without being notified about something!

Appendix 8 – IoT search engine knowledge

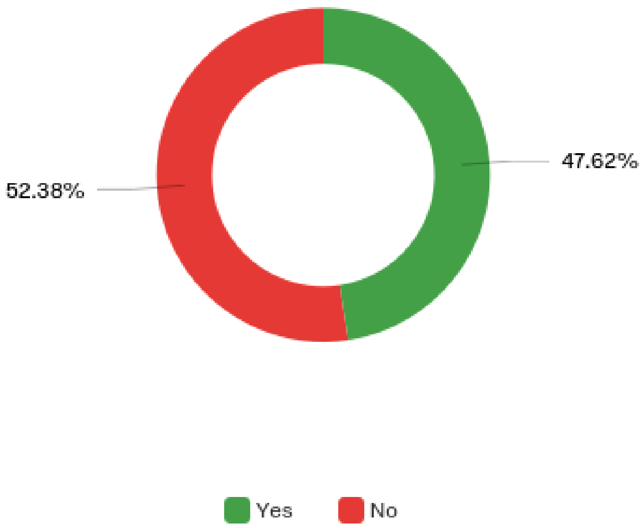
Q6.1 - Similarly to the search engines that are used to find content in the internet (e.g. Google, Yahoo!, Bing etc), there are Internet of Things search engines (e.g. Shodan, Thingful) that are used to find smart devices connected to the internet and provide relevant data to the user. Did you know about the existence of such search engines?



#	Answer	%	Count
1	Yes	30.43%	21
2	No	69.57%	48
	Total	100%	69

Appendix 9 – IoT search engine usage

Q6.2 - Have you ever used an Internet of Things search engine?



#	Answer	%	Count
1	Yes	47.62%	10
2	No	52.38%	11
	Total	100%	21

Q6.3 - Which was the last IoT search engine that you have used?

Google

shodan

Na

Shodan

shodan

Google

Shodan

shodan

Q6.4 - What term(s) did you search for? Please list up to 3 and press 'enter' after each term.

Aitken McKenzie builders

don't remember, it was only a try

Na

cameras fs

Cyberphypsical systems

Currency change Time in Auckland 1980 Kuwait stock crisis

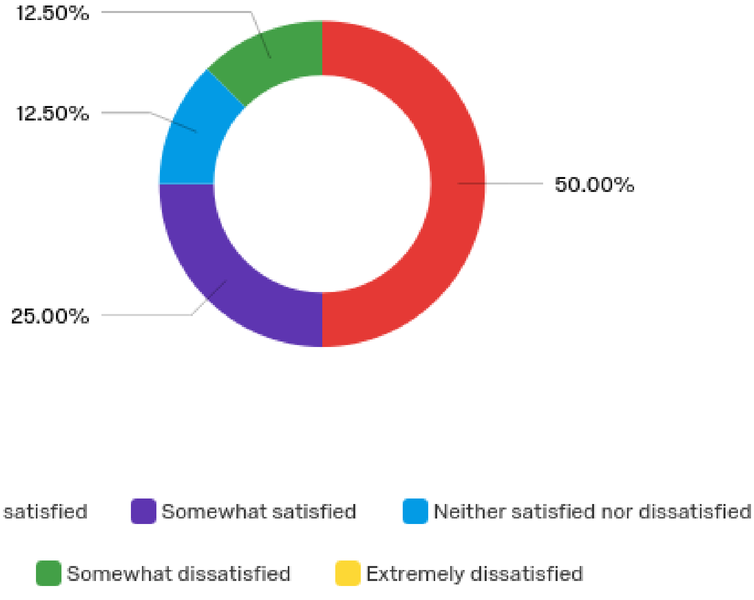
idrac ip:130.89.0.0/16 port:9200 ip:130.89.0.0/16 elasticsearch country:nl

security privacy relevance

Appendix 10 – IoT search engine satisfaction

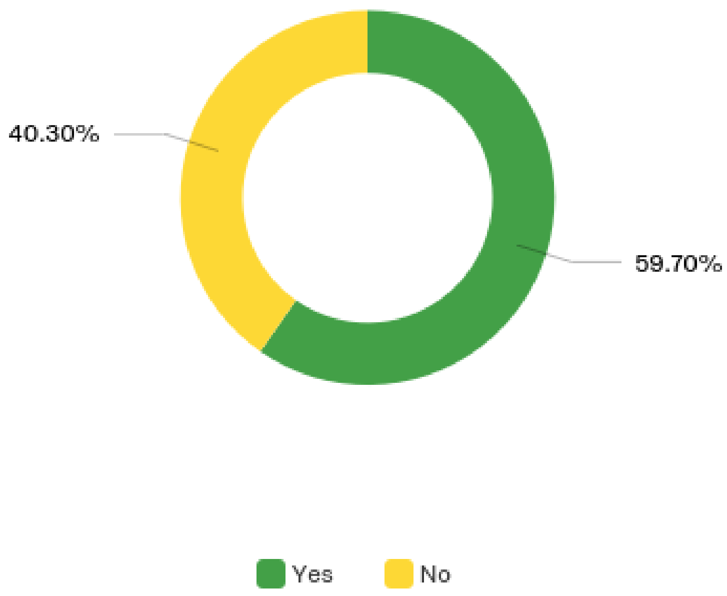
Q6.5 - How satisfied were you with the information retrieved?

#	Answer	%	Count
1	Extremely satisfied	50.00%	4
2	Somewhat satisfied	25.00%	2
3	Neither satisfied nor dissatisfied	12.50%	1
4	Somewhat dissatisfied	12.50%	1
5	Extremely dissatisfied	0.00%	0
	Total	100%	8



Appendix 11 – Smart devices usage

Q7.1 - A smart device is a device that can be remotely controlled and/or monitored when connected to the Internet or a local network. E.g. Smart thermostat, smart lights, IP camera, wearable device etc. Have you ever used such a device?



#	Answer	%	Count
1	Yes	59.70%	40
2	No	40.30%	27
	Total	100%	67

Q7.2 - What device(s) was/were that? Please list up to 3 that you have used.

Smart thermostat
Smart lights, Nest thermostat, Tesla
Smart TV , Smart lights , Smart watch
Thermostat
Google Home, Apple Watch
smart lights, google assistant
Smart thermostat in my family home
Smart thermostat
Camera
Amazon Alexa
Camera
Home weather station , Z-Wave mesh network with a 'panic button' for an elderly relative, SkyQ mesh network for multi-room reception and control of satellite TV
CNC mill, CNC lathe
Thermostat
Smart room control (in a hotel), Fitbit
WeVibe vibrator
Ip camera, Thermostat, Wearable
Nest thermostat
relays
Smart bulbs, Ip cameras
Cleaning robot, Baby monitor
Smart Music Player
Thermostat, cameras, audio sensor
Nest, hue, tv
desktop monitored by smartphone
Lights, ip cameras, air quality sensors
Apple watch, smart bulb, climote for home heating
Bluetooth detectors of the A1 and A68 trunk roads
Rpi, sensors, smart metering systems
IP camera in my apartment in order to check who is at the home. The thermostat in my apartment.
Wearable device, IP camera
Camera, Remote controlled car heating system, smart lights
wearable
Fitbit
Smartwatch
Internet Camera
In bioengineering we use weable devices often, accelerometers, emg these transmit data wirelessly to a computer

Appendix 12 – Online survey: Likert-type data

Home	Not at all important	Slightly important	Moderately important	Very important	Extremely important	Total Count	Sum
Camera (security camera, baby monitor etc.)	1	3	10	27	19	**60**	**240**
Smart lights	5	17	11	17	9	59	185
Smart thermostat	2	12	10	25	9	58	201
Virtual Home Assistant (Alexa, google home etc.)	8	11	14	12	9	54	165
Sensors for room temperature, humidity, air quality etc.	1	3	14	29	11	58	220
Other (Please specify)	5	3	9	4	5	26	79
Environmental	Not at all important	Slightly important	Moderately important	Very important	Extremely important	Total Count	Sum
UV sensor	3	3	10	19	20	55	215
Weather station	1	4	12	25	17	**59**	230
Air quality sensor	2	2	7	19	26	56	233
Water quality sensor	2	2	4	21	29	58	**247**
Other (Please specify)	5	0	11	4	4	24	74
Smart cities	Not at all important	Slightly important	Moderately important	Very important	Extremely important	Total Count	Sum
Parking availability sensor	2	4	8	21	21	56	223
Traffic camera	1	5	6	24	23	**59**	**240**
Charging point	2	4	14	24	10	54	198
Noise sensor	4	9	18	18	6	55	178
Other (Please specify)	7	0	13	3	1	24	63

Appendix 13 – Online survey: context sense responses

air quality	ship	radiation
air conditioning, fans, fridge	airport	background radiation sensor, UV sensor, RF/microwave detector
air filter, air cleaner	camera, remote steering, engine control, other mashines control/check	Computed Tomography Scanner
air freshener, air condition, air quality detector	Compas	Geiger counter
air quality alert on mobile phone	Engine room	Geiger meter
Air quality monitor	GPS	Gieger counter
Air quality sensor	gps	level,aiquality meter
air quality sensor	humidity detector, temperature detector, balance checker	measurement device
Air sensor, Weather sensor	lights, gps, smartphone	oven, bulbs, monitors
CO2 polution,NoX concentration	navigation aids, weather station	radiation detector, radiation warning, hospital-police auto informer
CO2 sensor	no idea	radiation level alert
heater, fan	radar	radiation measuring device
pollution monitor, weather station	Radar, sextant, sonar	radiation monitor
polution measuring device	radars	Radiation sensor
purifier	radars	radiation sensor
sensor	Satellite	radiation sensor
Sensor	satellite, antenna	radiation suit, geiger meter
sensor, filter	Sea Traffic sensor	radioactivity sensor
sensor, wireless router, storage devices	ship schedule,	radioactivity sensors
sensors	Sonar	sensor, warm detection, wireless router
sensors	sonar, filter, submarine	Thermostat
Smart phone	toy	UV sensor
weather station	weather,load	

earthquake	gamma	weather	shark
depth measure, level measuer	Detector	humidity,rain,temperature	Aqualung
Earth activity sensor	gamma radiation monitor	instruments used in weather stations	cage, camera, scubba diving equipment
earthquake alert, earthquake prediction,	gamma radiation sensors	Phone	gps
earthquake forecasting devices	Geiger counter	Satellite, weathet Station	GPS, sensors
earthquake prediction devices	measurement device	sensor	radiation system, wireless router, satellite
earthquake sensor	radiation detector	smart sensors	Satellite
Radar	radiation sensor	solar panels, antenna, anemometer	Sea Life Guard Sensor
seismograph	radiation sensor	Stevenson screen, thermometer, barometer	Shark detection device
seismograph	radiation station	sun sensor, weather station	shark detector, menacing Cello music detector!
seismographs	radiation suit, spectrometer,	thermometer	shark population monitor
sensor, computing server, wireless router	sensor	thermometer	ship
Siesmometer	signal strength,signal power	Thermometer	smart metering
Siesmometer	smart sensors, monitors	thermometer, barometer, forecaster	sonar
smartphone, bulbs, smart sensors		Thermostat	sonar , alert , danger
thermostat		Thermostat, hygrometer	sonic/ sound wave measurement device
vibration sensor		weather prediction	toy
voice detector, hospital-police auto informer,		Weather sensor	Tracker
weather station		weather station	vacuum cleaner
		weather station	
		wind speed sensor, thermostat	

temperature	**energy**	**aircraft**
high, UV detector, air condition	battery sensor, Smart Lights device	Air Traffic sensor
Phone	charging points, gas station	aircraft schedule, aircraft route report
sensor	consumpiton,smart metering	airport
Sensor	electricity measurement devic	airquality,fuel consumption
smart sensors, air conditioning	energy monitor	Binoculars
temperature control, temperature monitor	energy use monitor	Engine room
Thermometer	ergo-meter, smart utility meter	engine, black box, auto-pilot
thermometer	heater, socket	GPS
thermometer	network devices, sensors	gps,
thermometer,	Sensor	Radar
Thermometer, AC, humidifier, radiator	smart energy meter	radar
thermometer, heating, air condition	smart lamps	Radar
thermometer, pyrometer	smart meters,	radars
thermometer,uv meter	smart sensors	remote steering, engine control, other mashines control/check
thermometers	Thermometer, thermostat	Satellite
thermometre	Thermostat	satellite, network devices, sensors
thermostat	Turbine, voltmeter	System health monitoring unit
thermostat	wind turbine, solar panel, geothermal steam turbines	too many to pick just three!
thermostat		toy
thermostat		
thermostat		
Weather sensor, thermostat		
weather station		
weather station		
Weather station, thermometer		

Appendix 14 – Online survey: Context sense thematic analysis

Air quality	
Theme	**Count**
Air quality sensor/monitor	15
Pollution sensor	5
Weather station	3
Control/act	11
Alert/mobile phone	2

Ship	
Theme	**Count**
Geo-locating device	14
Control	4
Conditions	4
weather	2
other	4

Radiation	
Theme	**Count**
Radiation sensor/meter	14
Alert/warning	3
UV	2
Other	9

Earthquake	
Theme	**Count**
Seismograph/seismometer	7
Prediction	4
Alert	3
Other	9

Gamma	
Theme	**Count**
Radiation sensor/counter	13
Other	3

Weather	
Theme	**Count**
Weather station	6
Thermometer	5
Thermostat	3
Barometer	2
Predictor	2
Wind speed sensor	2
Other	10

Shark	
Theme	**Count**
Sonar/detector	8
Geo-location	3
Other	16

Temperature	
Theme	**Count**
Thermometer	15
Temperature Control	8
Weather	4
UV	2
Other	1

Energy	
Theme	**Count**
Monitor	12
Lights	2
control	3
Other	5

Aircraft	
Theme	**Count**
Geo-location	10
Aircraft status	7
Other	5

Appendix 15 – Focus group: Consent form

Consent Form

Name of department: Computer and Information Sciences
Title of the study: User-preferable features in an Internet of Things search engine.

- I confirm that I have read and understood the information sheet for the above project and the researcher has answered any queries to my satisfaction.
- I understand that my participation is voluntary and that I am free to withdraw from the project at any time, up to the point of completion, without having to give a reason and without any consequences. If I exercise my right to withdraw and I don't want my data to be used, any data which have been collected from me will be destroyed.
- I understand that I can withdraw from the study any personal data (i.e. data which identify me personally) at any time.
- I understand that anonymised data (i.e. .data which do not identify me personally) cannot be withdrawn once they have been included in the study.
- I understand that any information recorded in the investigation will remain confidential and no information that identifies me will be made publicly available.
- I consent to being a participant in the project

(PRINT NAME)	
Signature of Participant:	Date:

Appendix 16 – Focus group: Participant information sheet

Participant Information Sheet

Name of department: Computer and Information Sciences
Title of the study: User-preferable features in an Internet of Things search engine.

Introduction
Hello everyone and thank you for coming. I am Dimitrios Markogiannis and I am currently studying the postgraduate programme Msc Information Management at the University of Strathclyde. Email address: dimitrios.markogiannis.2016@uni.strath.ac.uk

What is the purpose of this investigation?
The aim of this investigation is to examine what features the users want in an Internet of Things (IoT) search engine so that they would be encouraged to use this kind of search engine and might lead to a greater and faster adoption of the IoT paradigm.

Do you have to take part?
Participation in this research is voluntary and you can refuse to participate or withdraw any time you want.

What will you do in the project?
You will be asked to draw your preferred user interface of a search engine after hypothetically use a given device as a search tem. After that, the two most popular IoT search engines will be demonstrated and I would like you to evaluate them and provide feedback on likes and dislikes. At the following stage some existing features of different search engines will be described and I would like you to rank them from most important to least important. Finally, you would be asked about recommendations for possible improvements.

Why have you been invited to take part?
You have been invited to provide your opinion on the subject of study.

What are the potential risks to you in taking part?
There are no risks in taking part.

What happens to the information in the project?
Your information will be kept confidential, and all data will be retained by the Information Services Directorate, University of Strathclyde on final completion of the dissertation.

The University of Strathclyde is registered with the Information Commissioner's Office who implements the Data Protection Act 1998. All personal data on participants will be processed in accordance with the provisions of the Data Protection Act 1998.

Thank you for reading this information – please ask any questions if you are unsure about what is written here.

What happens next?
If you are happy to participate you need to sign the consent form that is provided. The consent form is the only means of identification and proof that you participated in this focus group. Hence, it will be kept confidential in a separate location and it will not be included in the project report.

The results of this study will be published in the final report of my dissertation.

Researcher contact details:
Dimitrios Markogiannis, postgraduate student
dimitrios.markogiannis.2016@uni.strath.ac.uk

Supervisor contact details:

Dr Martin Halvey

Course Director MSc/PgDip Information Management

Telephone: +44 (0)141 548 3595
Email: martin.halvey@strath.ac.uk

Appendix 17 – Focus groups: Presearch questionnaires collected

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 26

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 27

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 29

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 24

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 23

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 39

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 22

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 33

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

Male Female Other / Prefer not to say

Age: 27

Have you ever used a search engine specifically designed for smart devices? (Please circle)

Yes No

Presearch questionnaire

Gender: (Please circle)

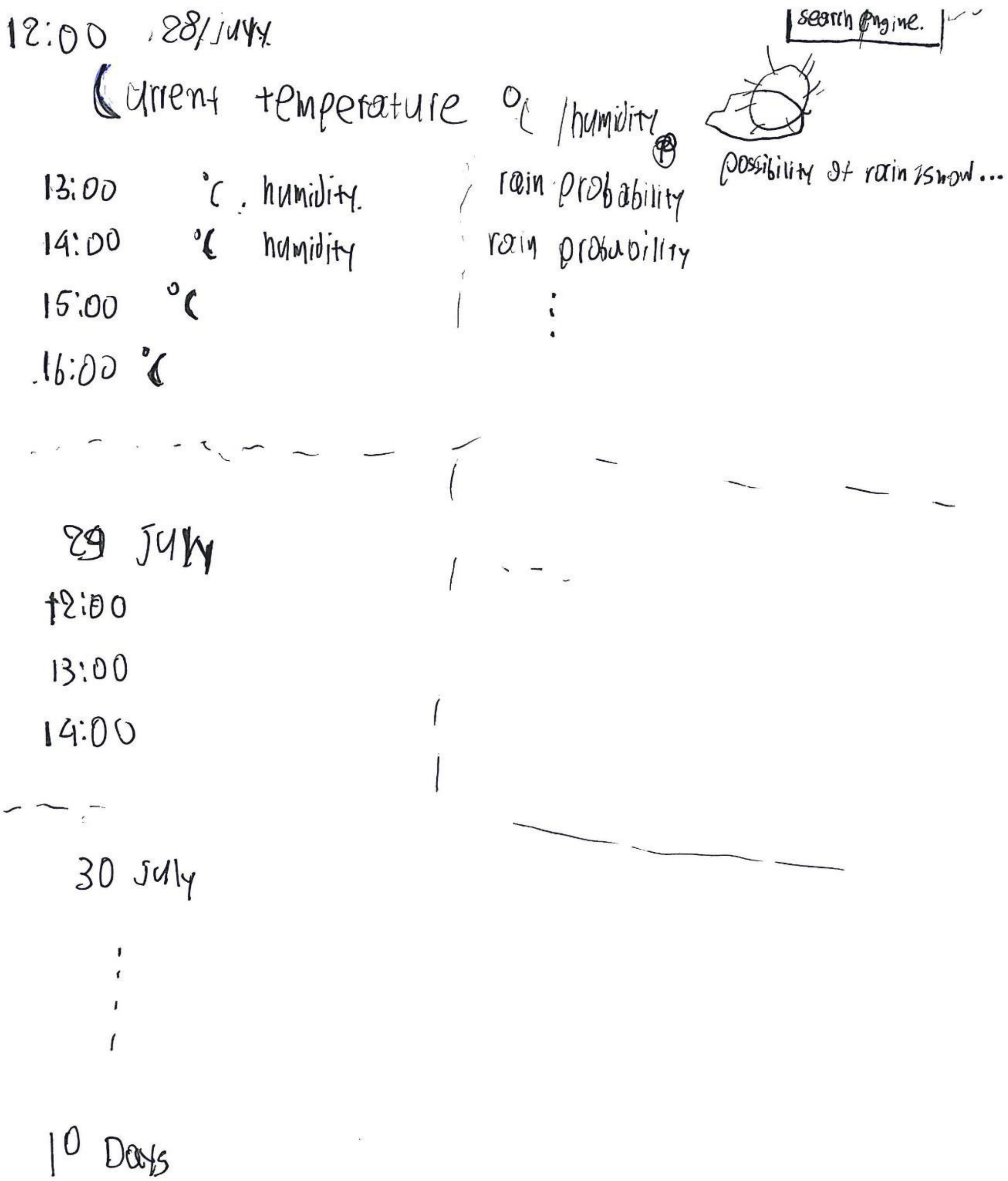
Male Female Other / Prefer not to say

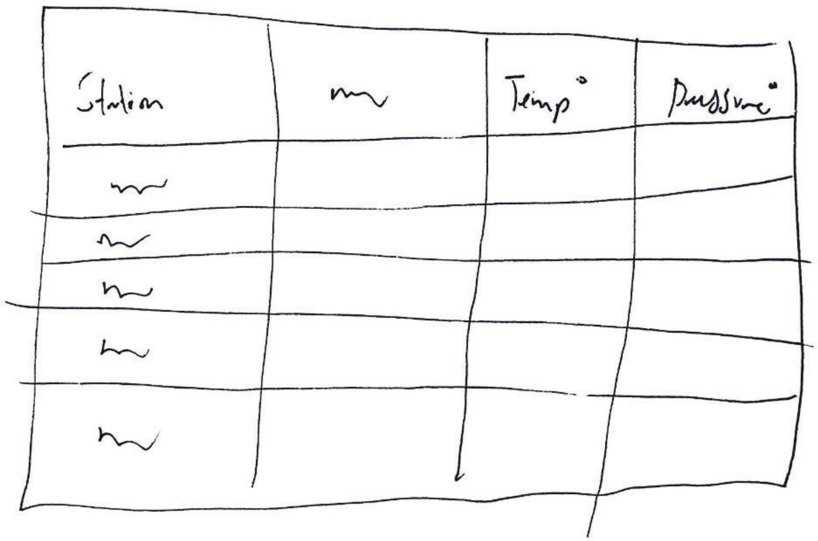
Age: 24

Have you ever used a search engine specifically designed for smart devices? (Please circle)

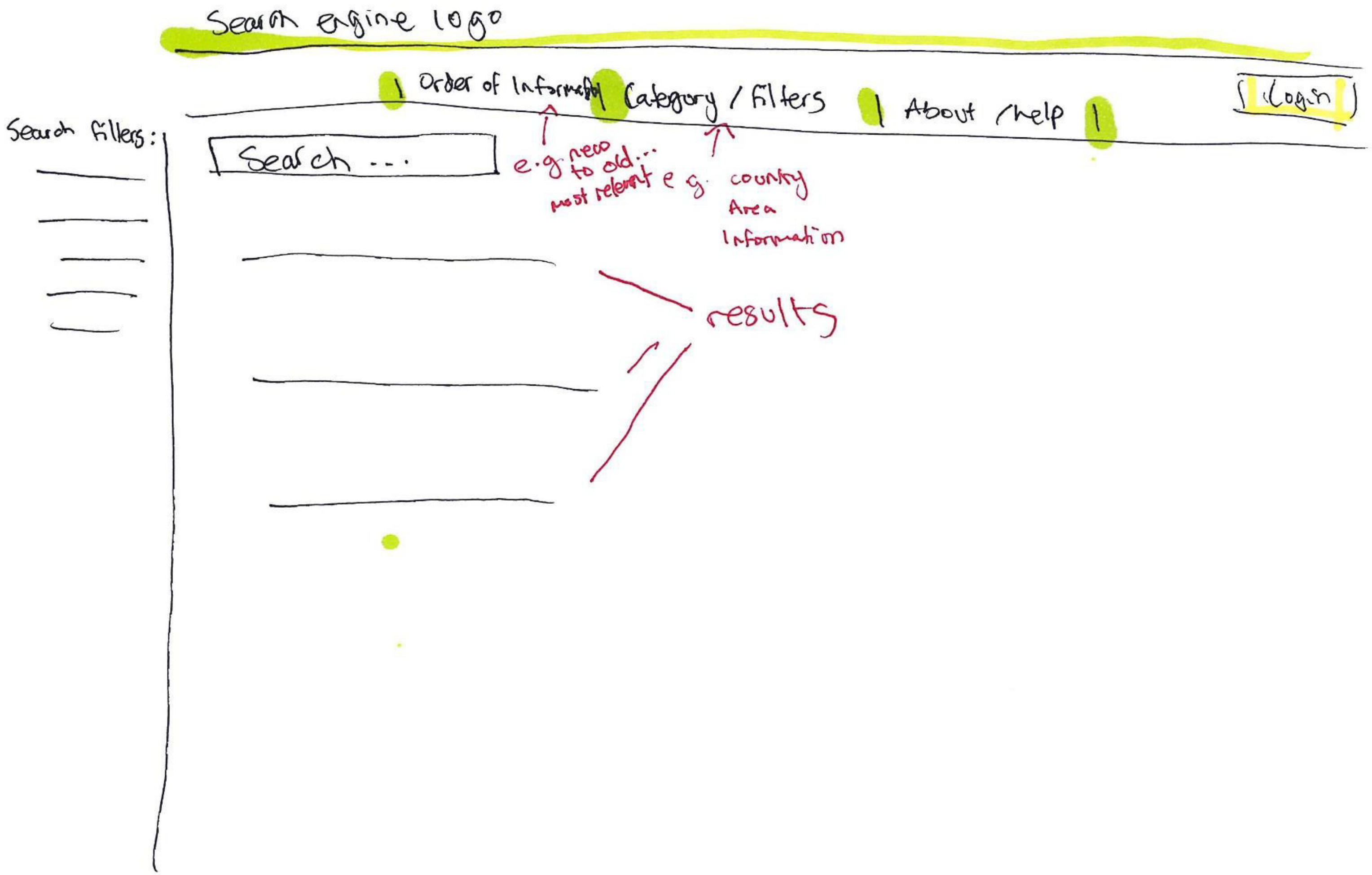
Yes No

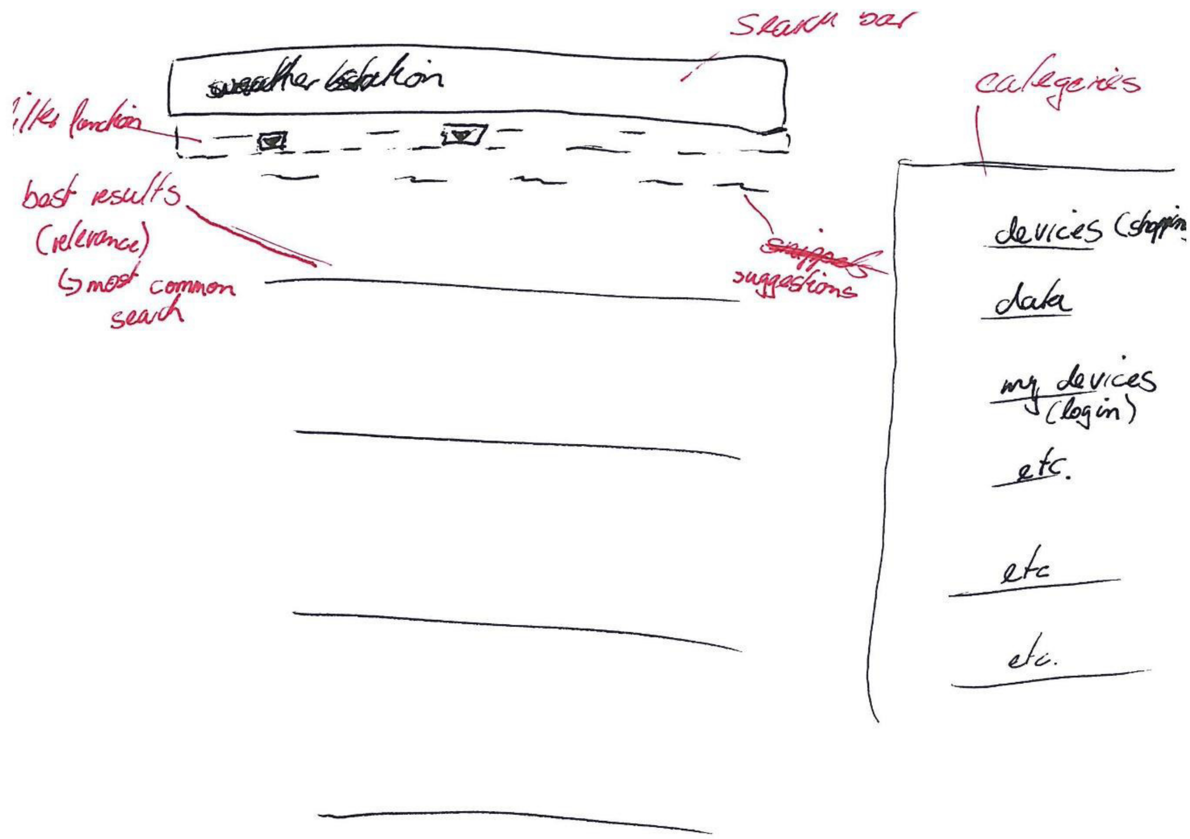
Appendix 18 – Focus group: Drawings





I'd like an API!


Search engine logo
Order of Information
Category / filters
About / help
Login
Search filters:
Search ...
e.g. new to old... most relevant
e.g. country
Area
Information
results



Search bar
weather station
categories
best results
(relevance)
most common
search
suggestions
devices (shoppin
data
my devices
(login)
etc.
etc
etc.

Country

language bottom

	°C	high	low
- Temperature.	___	___	___
- Humidity	___	___	___
- Rain	___	___	___
- Areas dry	___	___	___

Different cities in the country. temperature

City	city	city	city.
__°C	___°C	___°C	___°C

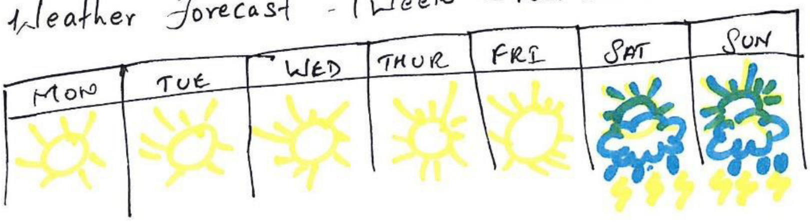
Catigories for Activities to do at different weather

~~Hiking~~ ______ ______ ______

Weather Station
Today
Date:
- Sunny
- cloudy
00 -
24
Hours
Temperatur / Wind speed
location
list of possibilities
week days

Search: Weather Station

- Weather Forecast - 1 Week <Your Location - Glasgow>



- UV ~~range~~
- Weather Forecast - ~~[illegible]~~ .Glasgow, Scotland <Different Sites).
- Travel disruptions/warnings
- Weather Forecast - London.

weather station

- My location, data about

- pick a location.

USA Europe Asia Africa Oceania North/South Pole.

- pick a device.

_____ _____ _____ _____ _____

- webpage information ~~what is the weather~~ about weather stations.
- webpage ~~must~~ most visited
- webpage most visited.

Search Result   New Search   Related Searches

Weather Station Glasgow

Glasgow Buchanan Str.

Summary Data Today

Temperature —
Humidity —
Wind —

☐ Download to Excel
☐ Print

Forecast Data for
-> following day
-> following week

=> see past data collection library
-> see further data tables on weather in Glasgow

Further topics
=> see possibilities for growing crops / farmers
=> danger of getting a sunburn
=>

People looked for this station also looked for
-> weather station Edinburgh
-> weather station Loch Lomond
->

Related Topics
- Station is powered by Company XY - see further topics of this company

Weather station results.

What data can be collected

-
-
-
-

What sensors are used to collect the data

-
-
-
-

What format is the data in, and how can it be sent / transmitted to other devices

-
-

Appendix 19 – Focus group: Completed ranking forms

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	3
Personalised search	6
Sorting	5
Filters	4
Aggregated search	1
Facets	2

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	4
Personalised search	5
Sorting	3
Filters	1
Aggregated search	~~3~~6
Facets	2

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	2
Personalised search	3
Sorting	6
Filters	5
Aggregated search	1
Facets	4

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	4
Personalised search	6
Sorting	3
Filters	2
Aggregated search	5
Facets	1

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	2
Personalised search	4
Sorting	6
Filters	5
Aggregated search	1
Facets	3

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	5
Personalised search	6
Sorting	3
Filters	2
Aggregated search	4
Facets	1

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	1
Personalised search	6
Sorting	3
Filters	2
Aggregated search	5
Facets	4

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	2
Personalised search	1
Sorting	3
Filters	5
Aggregated search	6
Facets	4

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	4
Personalised search	3
Sorting	5
Filters	6
Aggregated search	1
Facets	2

Please rank the features below. Consider 1 as the most important and 6 the least important.

Feature	Position
Text snippets	2
Personalised search	5
Sorting	4
Filters	3
Aggregated search	1
Facets	6

Appendix 20 – Focus group: Transcript

Focus group 1

Shodan

Participant 1

-Hahaha

-Not for us, the common persons.

-We need the data behind the weather station. What data you can retrieve from that weather station.

Participant 2

No comment.

Participant 3

-How exactly are you using that?

Participant 4

- Obviously you want to use the data. Wind, humidity temperature.
- Suggestion: Shortcut links to the database,
- Suggestion: some analysis. Average temperature, average rain.
- What others looked for, related topics.
- Shodan doesn't look good yet.
- Suggestion for analysis data.
- Limited options.

Thingful

Participant 1

- This is better

Participant 2

- Yes, it is better. More user friendly. Has way more relevant information to us common people.

Participant 3

No comment.

Participant 4

No comment

Focus group 2

Shodan

Paticipant 1

- Interesting because it gives the data sources. It illustrates the real big problem. You've got the simplicity of the interface, and it is quite simple to input search terms and stuff, but then the technical level of the information you're getting out is quite high.

Participant 2

- Makes me totally not happy. I cannot relate to those numbers.

Participant 3

No comment

Participant 4

- What do the numbers mean?

- Suggestion: Filter for public or private. Accessible or not accessible.

- I guess not everyone would be looking for such information to be very specific.

- People would know how to search for it.

Thingful

Participant 1

- It looks really cool. If I wanted to actually see information from a weather station I would probably want a big table of data. There is not much I can do by just clicking different weather stations. Weather data would be interesting to see. For example see data about a fridge and see what extra energy should be consumed on hot days.

- Apart from just getting an overview on what there is out there, I can't find something useful.

- It would be nice to have visualisation of the data.

Participant 2

- Nice user interface. But it's more playing around than getting something useful out of that. Better than Shodan but I can't relate to do anything useful with this data at the moment.

- The small number of devices means it's not mature; not complete. If as a user you know that the search engine that you use it doesn't show all the results, it doesn't make you satisfied.

Participant 3

- I think it's used by specialists.

Participant 4

- It's visually pleasing.

Focus group 3

Shodan

Participant 1

- If you are looking for it might be useful.

- I like the way it's set up. Nice and clear this is this device. It's very clear which one you are clicking on.

Participant 2

- Why private devices exist there if we can't access them?

- I prefer to see the details directly, I don't want to click to see the information.

Thingful

Participant 1

- I like that you can see where things are on the map. For weather station and public devices you can see how close are to each other. But for personal devices is very dangerous. I prefer the other way the information was displayed. In nice clean…(Incomprehensible). With that, you have no idea what it is without clicking on it.

Participant 2

No comment.

Appendix 21 – Focus group: Presentation

Internet of Things
(IoT)

A network of devices, equipped with advanced sensors, that are able to collect and exchange data.

IoT search engine

- An "IoT search engine" is a search engine that enables the collection of data from heterogeneous sources such as the smart devices.

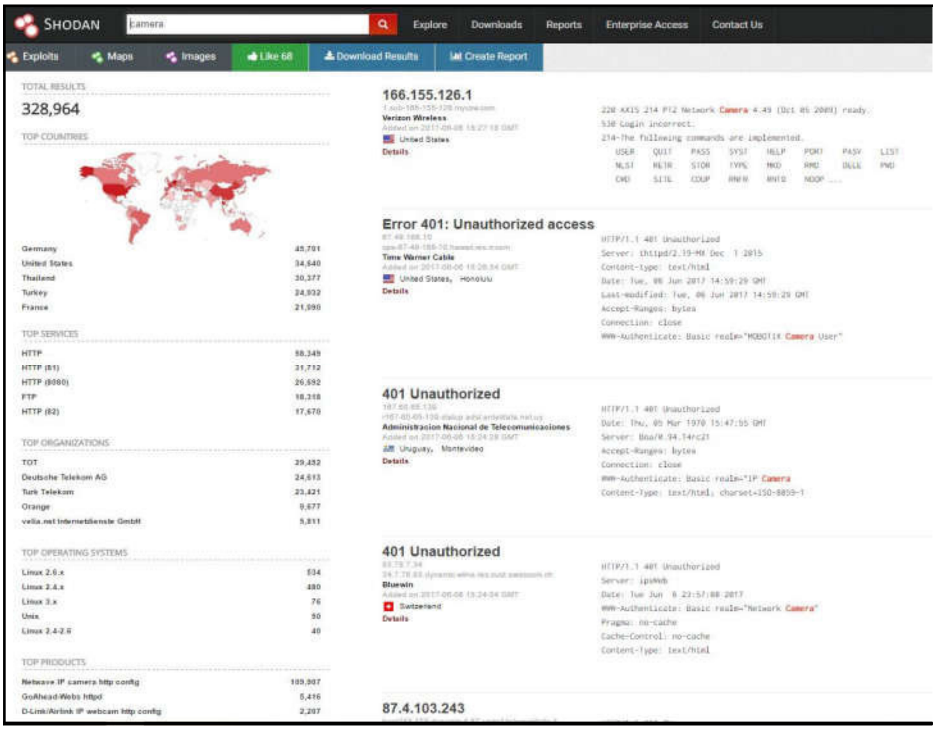
Use case

The management of the company you work for, has decided to exploit new technologies and use the information that can be retrieved from smart devices and IoT, to gain competitive advantage. As part of forming the company's strategy, you have been asked to find and document what data can be retrieved from a smart weather station.

To complete the task you use an IoT search engine.

Please draw how you would prefer, the retrieved information to be demonstrated in the result-page.

Existing IoT Search engines



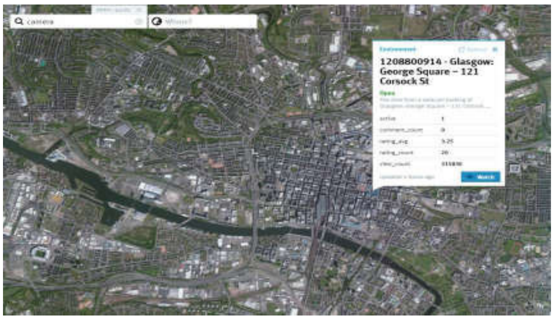
Shodan search results for ‘camera’

Thingful search results for ‘camera’

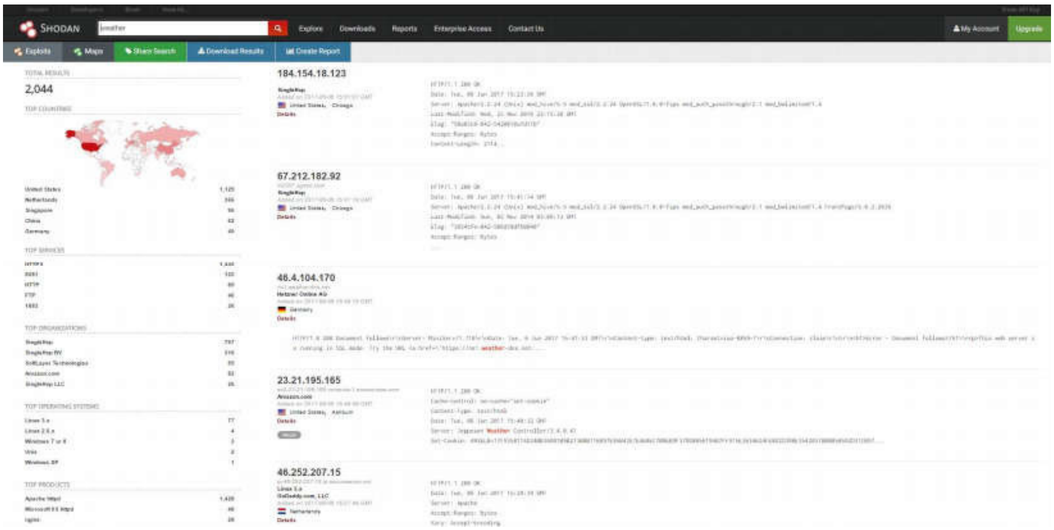
Thingful search results for ‘camera’

Before selection

After selection



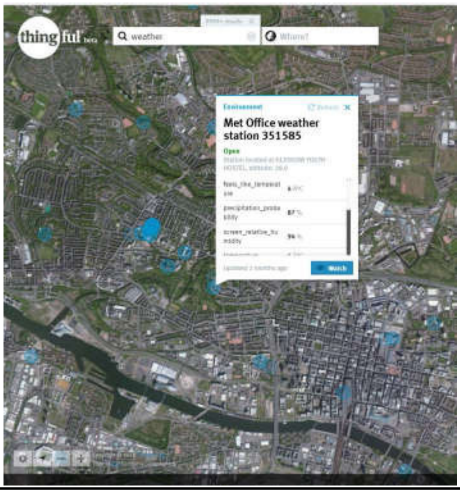
Shodan search results for ‘weather’



Thingful search results for 'weather'

Before selection

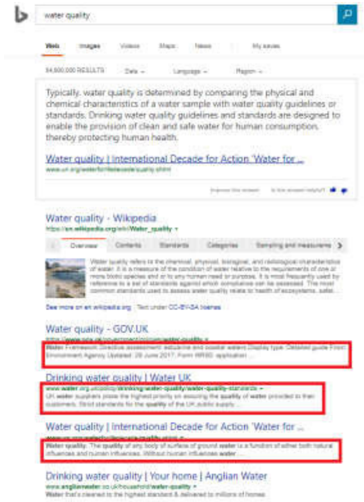
After selection



Search engine features

Search engine features

Text Snippets

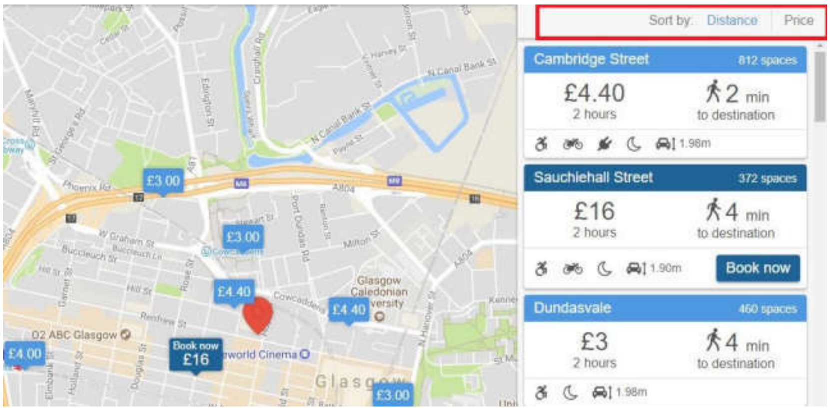


Search engine features

Personalised search

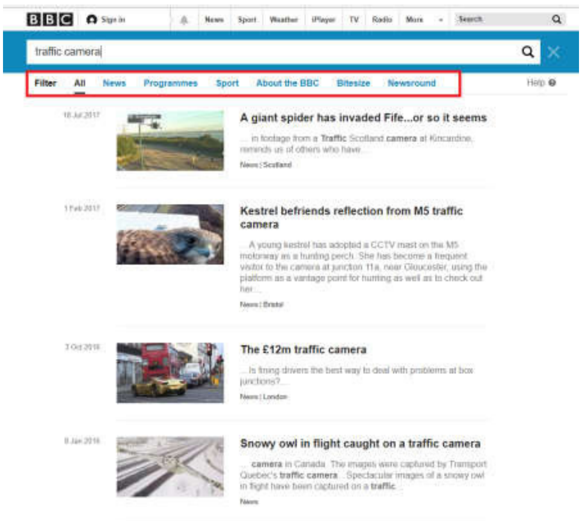
Search engine features

Sorting



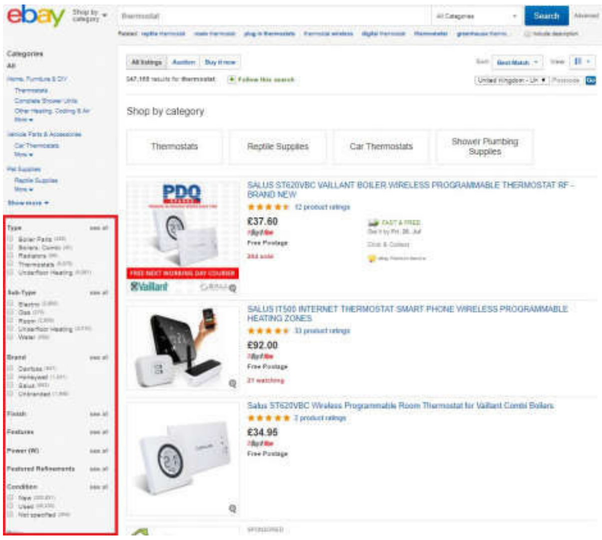
Search engine features

Filters



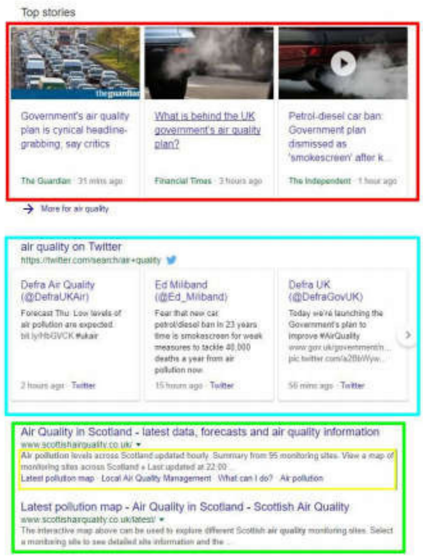
Search engine features

Facets



Search engine features

Aggregated search (Rich content)



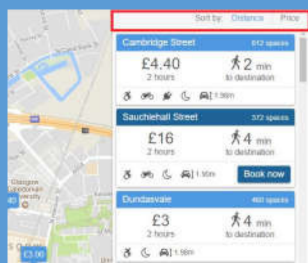
Ranking the features

Please rank the features in the form. Consider 1 as the most important and 6 the least important.

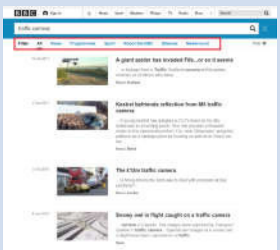
Text snippets

Personalised search

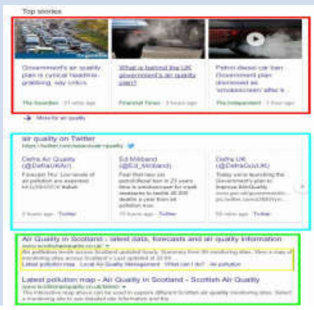
Sorting



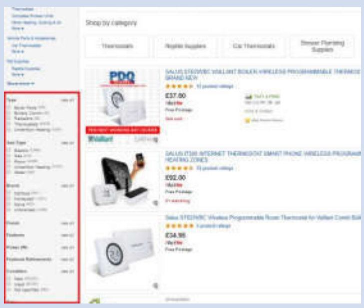
Filters



Aggregated search



Facets



Thank you!