Report on the 1^{st} International Workshop on Information Access in Smart Cities (i-ASC 2014)

M-Dyaa Albakour[1] Craig Macdonald[1] Iadh Ounis[1]
Charles L. A. Clarke[2]
Veli Bicer[3]
[1] University of Glasgow, Glasgow, UK
{*dyaa.albakour,craig.macdonald,iadh.ounis*}*@glasgow.ac.uk*
[2] University of Waterloo, Waterloo, Canada
claclark@plg.uwaterloo.ca
[3] IBM Research Lab, Dublin, Ireland
velibice@ie.ibm.com

Abstract

Modern cities are increasingly becoming smart where a digital knowledge infrastructure is deployed by local authorities (e.g. City councils and municipalities) to better serve the information needs of their citizens, and to ensure the sustainability and efficient use of power and resources. This knowledge infrastructure consists of a wide range of systems from low-level physical sensors to advanced sensing devices through social sensors. The i-ASC 2014 workshop was the first international event, within the Information Retrieval (IR) community, that is dedicated to research on smart/future cities. In particular, the workshop was a venue for research on digesting the city's data streams and knowledge databases in order to serve the information needs of citizens and support decision making for local authorities. Possible use cases include helping tourists to find interesting places to go or activities to do while visiting a city, or assisting journalists in reporting local incidents. Indeed, the workshop was intended to foster the development of new information access and retrieval models that can harness effectively and efficiently the large number of heterogeneous big data streams in a city to provide a new generation of information services. The workshop was well attended, where more than 45 participants were officially registered. It featured two keynote talks from industry (IBM and Waag Society) and two invited talks from academia (Pisa and Edinburgh). In addition, seven refereed papers were presented before breakout groups considered questions and issues identified from a panel discussion.

1 Introduction

Cities of the 21^{st} century do not only have the physical infrastructure of roads, buildings, and power networks, but also have the knowledge infrastructure represented with heterogeneous

systems and big data platforms. These systems vary from low-level sensing devices including environmental sensors and CCTV cameras, to public databases, and social network streams. On the other hand, a recent study [9] has identified a variety of emerging information needs that citizens often have in their public urban spaces. These local information needs may be complex and are not necessarily served by existing systems such as web search engines. These include obtaining real-time information about events in the city or finding free parking spaces. The aforementioned knowledge infrastructure opens up opportunities for cities to become smarter and serve the emerging information needs of their citizens by harnessing the vast amount of diverse information stemming from the various systems within the city.

The i-ASC 2014 workshop was an opportunity for researchers from the IR community, and other related communities, to discuss the emerging research topic of information access in smart cities, identifying its unique challenges, opportunities and future directions. In particular, we identified four main research themes that we aimed to cover in i-ASC:

Searching smart cities: Searching smart cities has unique characteristics and may not be necessarily served by existing IR models and techniques. A recent tutorial at SIGIR has highlighted some of these characteristics. [1] Firstly, the information needs of citizens can be complex and require the consolidation of a wide variety of city data sources [9]. Secondly, the relevance of results may depend on a number of dimensions such as geospatial, temporal and social dimensions. Furthermore, real-time search plays an important role in such environments, where it is vital to obtain the freshest results possible. New retrieval models are emerging to tackle these unique characteristics [1].

City data: There is a wealth of data sources that can be acquired about modern cities. However, city data sources have unique characteristics. First of all, the heterogeneity of this data poses challenges for identifying the relevant sources of information for specific needs, e.g. from different repositories in the linked data cloud [11], or otherwise fusing multiple sources of information including sensor streams [4]. Also, this data can be noisy, messy and incomplete, which adds an additional burden of cleansing, capturing the anomalies, and consolidating multiple sources. Moreover, the volume of such data can be huge and requires an efficient computing infrastructure that scales to large amount of knowledge databases and data streams. The emerging big data platforms play a significant role in providing such infrastructures.

Context and Recommendation: Recently, the SWIRL 2012 workshop [3] has noted that future IR systems should be able to anticipate user information needs without explicit queries, i.e. zero queries. This is particularly suitable for mobile devices, which impose a physical limitation on the user interactions. The TREC contextual suggestion track [5] introduced a typical example of a zero-query task in smart cities, where the objective is to recommend to a user interesting activities to do in a city based on contextual information of previous preferences. Developing zero-query access to information requires effective modelling and representation of contextual information about the users and the environment in order to both *anticipate* the user needs and *personalise* the results.

HCI and Applications: Providing access to information in smart cities requires suitable user interaction paradigms. The interaction paradigm may have impact on the design of the underlying retrieval models and the user interfaces. The design of user interfaces should support the exploration and discovery of city data on emerging mobile devices to make them even more accessible. Moreover, there is a long a list of domains, which can benefit from

[1] http://www.dublinked.ie/?q=searchinginthecityofknowledgetutorial

applications that employ such interfaces. This includes transportation [4], tourism (recommendation platforms, access to heritage information), entertainment (live news), security and surveillance (crime prediction and prevention), crisis management [14], and urban planning.

2 Programme Committee

The formed programme committee was balanced between industry and academia, as well as a coverage of expertise in the themes identified above. The members in alphabetical order are:
Omar Alonso (Microsoft Bing)
Fernando Diaz (Microsoft Research)
Cathal Gurrin (Dublin City University)
Daqing He (University of Pittsburgh)
Jaap Kamps (University of Amsterdam)
Freddy Lecue (IBM Research)
Franco Maria Nardini (ISTI CNR)
Raffaele Perego (ISTI CNR)
Paul Thomas (CSIRO)
Suzan Verberne (Radboud University Nijmegen)

3 Workshop Programme

The Workshop programme included four invited talks, and seven technical papers, which culminated in three breakout groups. The list of talks is the following:

Invited Talks:

- *Content, Connection and Context: From Data to Insight in Smarter Cities.* Pól Mac Aonghusa.
- *Smart Cities, Smart Citizens and the case for the CitySDK.* Frank Kresin.
- *Mining digital footprints for smart tourism.* Raffaele Perego.
- *Tourists in Smart Cities: Data mining for hidden treasures.* Jon Oberlander.

Accepted Technical Papers:

- *From Smart Cities to Smart Neighborhoods: Detecting Local Events from Social Media.* Yang Li and Alan Smeaton.
- *Modeling the Web of Things from an IR approach.* Cristyan Manta-Caro and Juan M. Fernndez-Luna.
- *The Influence of Indoor Spatial Context on User Information Behaviours.* Yongli Ren, Martin Tomko, Kevin Ong, Yuntian Brian Bai and Mark Sanderson
- *On Mining Mobile Users Monitoring Logs.* Dmitry Namiot.
- *Challenges in Recommending Venues within Smart Cities.* Romain Deveaud, M-Dyaa Albakour, Craig Macdonald and Iadh Ounis.
- *Smarter Cities, Safer Travels: Integrating Contextual Suggestion.* Adriel Dean-Hall and Jack Thomas.

Figure 1: A word cloud generated from the contents of the #iasc2014 tweets

- *"Search-the-City" - A versatile dashboard for searching and displaying Environment and User Generated Content in the context of the future Smart City.* Athanasios Moralis, George Perreas, Anastasios Glaros and Dimitrios Dres.

The full workshop proceedings and the presented slides are available to download from the workshop's website.[2] The programme was structured as four sessions where the four aforementioned themes were covered in the first three sessions by four invited talks, and short presentations of the accepted papers. The last session featured a panel to reflect on the talks and to identify the discussion topics for the breakout groups. To give the reader a flavour of the workshop's presentations and discussions, in Figure 1, we show a word cloud generated from the tweets posted on the workshop day with the hashtag #iasc2014. In the following, we summarise the workshop's activities.

3.1 Theme 1: Searching Smart Cities

After a short introduction from the organisers, the workshop kicked off with a session dedicated to searching smart cities. The session started with the invited talk of Pól Mac Aonghusa [12], who presented broad views of research in knowledge-driven urban systems and the challenges that face IR in these systems. These views are supported by the various real-world use cases that he has been involved in as a leading researcher in IBM Research, Dublin. Pól highlighted the issues of the heterogeneity of (Big) data in the city data ecosystem, where data sources can range from factual static data about locations and demographics and (near) real-time data of traffic and sensor observations, to sensitive data collected by national bodies, such as data about the city financial performance. Therefore, IR in such environments has the challenge of finding the right sources of data for a specific task, and also the challenge of semantically linking this data, which is currently neglected in smart city systems. With regards to the semantic linkage issue, Pól noted that a dramatic improvement of the effectiveness of an IR system in smart cities can be actually achieved by assisting data

[2] `http://dcs.gla.ac.uk/workshops/iASC2014`

providers in labelling their data. Finally, Pól touched upon the importance and the richness of context in these IR systems. Context exists in various forms in such environments. This includes, for example, the social context of citizens in integrated citizen care systems, which can be identified in the citizens' specific needs and their community among other aspects. Other forms of context include spatial and temporal context as in the spatial context for diagonising events from massive real-time streams in the Simplicity EU FP7 project.

Following the keynote talk, two papers were presented to discuss IR problems in smart cities. The first talk, presented by Alan Smeaton, considered the problem of identifying local events from the Twitter social network [10]. Using the city of Dublin as an example, Alan presented a method of partitioning the city into neighbourhoods based on the geographical occurrence of tweets. The problem of detecting events in these neighborhoods is then tackled by estimating the deviation from the regularity of tweeting in three different dimensions: (1) the time of tweets, (2) the exact location, and (3) the semantic consistency using a language modelling approach. Alan validated the detection of local events in these partitions from few examples, but he highlighted the issue of the unavailability of datasets to evaluate such IR tasks. Building such datasets is also challenging, since it is hard to judge some localised events in remote areas, e.g. who would know about slow traffic on M50 near the Blanchardstown exit on the morning of 5th March 2013?

Another IR problem was discussed in the second talk of Juan Fernandez-Luna [13]. Juan presented a theoretical data model for making the Internet-of-Things (IoT) searchable. In this model, the physical world is abstracted to contain two main types of elements: *things* representing tangible or intangible objects, and *spaces* where these objects exist. The proposed model use an XML representation to handle the variety of information about these objects and their spaces, such as their keyword descriptions, thier classifications and the real-time information from sensors attached to them. Such data can be then fed into search engines to answer novel types of information needs such as finding "*restaurants with a terrace that are located in the city center.*"

3.2 Theme 2: City Data

The second session covered the City Data theme and started with an invited talk by Frank Kresin from the Waag Society in Amsterdam [8]. Frank presented his views on designing smart city systems and the challenges of integrating smart city data. His views are particularly informed by his expertise through the involvements in various international innovation programmes such as "Apps for Europe", "City SDK", and "CineGrid Amsterdam". Frank started by reminding the audience with the importance of raising and deserving users' trust through data reciprocity and transparency. He argued that citizens should be the focus when designing smarter city systems, where they should be designed "with" people, not "for" or "against" them. This is the main principle followed when developing the City Service Development Kit (CitySDK) (http://www.citysdk.eu), which aims to deliver a set of bottom up harmonised APIs for software development to tackle a major challenge of smart city applications, namely the scalability across cities and countries. The challenge of scalability stems from the non-uniformity of data formats and APIs between cities. This is often amplified by the lack of documentation in these systems and the heterogeneity of data sources. The CitySDK Linked data API, within the CitySDK initiative, makes progress towards providing interoperable interfaces for open city data services by developing a uniform distribution API. Several applications (mobile apps) have been built on top of this platform in various cities

such as tourism applications, city dashboards, and smart parking apps.

Next, Mark Sanderson presented a technical paper on analysing large datasets of web access logs collected from wireless hot spots in a shopping mall [19]. In particular, Mark echoed Pól's discussion in the first invited talk about the importance of context, where he presented a study of how a spatial context in the studied indoor environment may influence the content that users may access on the web. Indeed, Mark has noted that organisations running commercial indoor spaces are looking for sources of user behaviour data, of which online retailers currently have access to, and use it to enhance their services and to increase their profits via advertising and recommendation. Mark concluded his talk with the main finding of the study, which suggests that the information access behaviour of users manifests a significant location-based bias when the common information behaviour is excluded. Following this, new research challenges arise on how to make use of the differences in information behaviour for delivering better services.

Following Mark's talk, Dmitry Namiot presented a technical paper on passive monitoring of mobile devices in wireless networks [16]. While in the previous talk, the aim was to investigate whether there is a correlation between the accessed content and the location of the user, the work presented here takes a content-agnostic approach and aims to identify groups of users and visit patterns from the raw access data. The model introduced has been tested in a café building to cluster groups of visitors.

3.3 Theme 3: Context and Recommendation

The Context and Recommendation theme started with an invited talk by Raffaele Perego [18]. Raffaele presented a number of approaches for mining social media streams and knowledge databases for the recommendation and personalisation of Point-of-Interests (POI)s to tourists. Raffaele started by surveying social media platforms that can be leveraged for POI recommendation and concluded that the Flickr photo sharing platform has the advantage over other social networks, such as Twitter and FourSquare, to be used in order to understand tourists behaviour since it is biased towards tourists. The challenge however is to cluster the Flickr photos and to match them with relevant POIs. Raffaele presented approaches that exploit the social tags available in Flickr and the visual content of the photos in Flickr to cluster photos and match them to POIs. Furthermore, Wikipedia categories were employed to devise patterns of tourists' behaviour by building trajectories of categories visited during a single journey. Building on these trajectories, Raffaele gave a demo of the TripBuilder system which can recommend trajectories (trip routes) to a user based on their personal interests and time constraints. This is done by solving an optimisation problem, which finds a subset of trajectories that maximises suitability to the user's interest, and fits in the time budget.

Following the invited talk, Romain Deveaud [7] presented a position paper discussing the challenges for venue recommendation systems. In particular, Romain highlighted the role that physical sensors, such as cameras, microphones and GPS sensors, can play in the future to provide relevant suggestions of venues to visit. Such sensors provide real-time data about the current status of the environment, which guide the recommendation process. For example, real-time traffic information can be used by a recommender system to recommend areas that can be reached easily. In addition, mining observations over a long period of time can help in forecasting future events, which can be used to guide venue recommendation.

The Context and Recommendation theme was closed with a talk by Adriel Dean-Hall [6] who looked at venue recommendation from the interesting new perspective of safety. Adriel

investigated how public crime data can be integrated into venue recommendation systems in order to consider safety information. Using crime data from the US, a safety score is estimated and integrated into the venue suggestions to provide contextual information to users. Although this can be useful, Adriel noted the implications that arise when naming a certain area as unsafe, as businesses and residents may not appreciate it.

3.4 Theme 4: HCI and Applications

The last talk in the workshop was an invited talk by Jon Oberlander, University of Edinburgh [17]. Jon has been involved in the SICSA Smart Tourism Programme,[3] which ran a number of projects over the last three years. In his talk, Jon gave a flavour of what tourism applications can do to exploit city data for (1) improving the personal experience of visitors in navigating a city, and (2) helping authorities in better managing their resources during times of peak demand. The dot.rural project is an example of future touristic applications that can leverage existing catalogues and knowledge databases for touristic recommendations. It uses observations about flowering plant species from the GBIF open data repository and textual data from plant databases, such as Flora Celtica, to recommend botanical POIs where certain plants flower at certain times. Jon presented the emergence of new channels and approaches to increase access to content and reach wider audience. Most notably, Jon presented a gamification approach, piloted in Edinburgh, to attract people into the city's museums by playing mobile games where the aim of the game is to collect historical objects (treasures) around the city.

Finally, in [15], Telesto technologies presented a visualisation framework for smart city applications. The framework is built upon the SMART EU FP7 project and aims to empower developers and city stakeholders to rapidly develop dashboards and end-user applications using customisable mashups that can present environmental and user-generated data in a meaningful way.

3.5 Panel and Breakout Groups

The last session of the workshop started with a panel discussion to reflect on the day and identify topics for the following breakout group discussion. Jon Oberlander, Mark Sanderson, Raffaele Perego and Charles Clarke were invited to the panel. First, each panellist gave an introduction about their background and their specific interests in smart cities research. The discussion was then driven by a number of questions raised by the panel moderator, Craig Macdonald, to elaborate on some of the challenges and research opportunities that surfaced from the various talks throughout the day. The discussion helped in identifying a number of topics that audience particularly showed interest in. As a result three breakout groups were formed:

- **City data**: the data breakout was chaired by Jon Oberlander and mainly considered how city data can be more open to be exploited in smart city applications and research. City data can exist in archived formats of databases or real-time streams from sensors. The main stakeholders in the process of sharing and consuming city data include commercial organisations, city authorities, scientists and citizens. In order to make city data more open, the main barriers behind not sharing city data should be understood.

[3] http://www.smarttourism.org/

The group suggested a couple of reasons that may prevent providers from sharing this data. The first one is the fear from the consequences of misusing this data, something that can outweigh the desire for change. The second reason is that data is often not ready to be shared (e.g. not clean) and requires some additional steps to be formatted, such that it supports linking and discoverability.

- **Privacy concerns**: the privacy group was chaired by Jack Thomas, and discussed mainly privacy concerns that may arise in smart city applications. Smart city projects are reliant on data from governments and corporations, who naturally wish to preserve the privacy for any data they could profit from, or data that reveals matters of corruption or misconduct. On the citizens' side, another privacy concern may arise from bringing multiple data sources together that can de-anonymise people unexpectedly. Finally, the group has noted the importance of engagement to tackle privacy concerns. Data and services should be built with a bottom-up approach by engaging with people whom we collect information on. While people do not like much of the surreptitious data collection about them or their environment, if the benefits of the services built with this data are explained to people, they are more likely to co-operate.
- **IR tasks in smart cities**: the tasks breakout group was chaired by Mark Sanderson and discussed examples of emerging IR tasks in smart cities that the IR community can work on, and the challenges of performing such tasks. One good example is the TREC Contextual suggestion track, which aims to provide recommendations of activities to do whilst visiting a city without having an explicit query (zero-query). The group discussed the common characteristics of such tasks. First, the main theme of these tasks is that their aim is to make city data more accessible. Second, they aim to help citizens improve their quality of life. An example of an IR task could be to find the best neighbourhood for a citizen to live in. The major challenge for the community is that often data sources are not necessarily publicly available. However, the IR community should aim to start with current city open data, such as the Dublin data presented by Pól in the first invited talk (Dublinked). Finally, following this discussion, the group and the participants of the workshop have shown great interest in developing new smart city tasks that can be studied in IR evaluation frameworks such as TREC and CLEF using open data.

4 Conclusions

After the breakout groups finished their discussion, a representative of each group presented their findings. Each of the groups contributed to our understanding of the way forward. For example, the data group identified possible avenues for encouraging city data providers to share data. The privacy group summarised the privacy matters in smart city applications and ways to handle them. Finally, the tasks group identified the properties of IR smart city tasks and put forward interesting exemplary tasks that the community can start working on. Following the success of this first i-ASC workshop, we aim to follow it up by another edition in a future, perhaps via more diverse, conference such as WSDM or WWW.

References

[1] M. Albakour, C. Macdonald, and I. Ounis. Identifying local events by using microblogs as social sensors. In *Proceedings of OAIR'13*, pages 173–180, 2013.

[2] M. Albakour, C. Macdonald, I. Ounis, V. Bicer, and C. L. A. Clarke (Editors). Proceedings of the Information Access in Smart Cities Workshop (i-ASC 2014). Amsterdam, the Netherlands.

[3] J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. *ACM SIGIR Forum*, volume 46, pages 2–32. ACM, 2012.

[4] E. M. Daly, F. Lécué, and V. Bicer. Westland row why so slow? fusing social media and linked data sources for understanding real-time traffic conditions. In *Proceedings of IUI '13*, pages 203–212, 2013.

[5] A. Dean-Hall, J. Kamps P. Thomas, E. Vorhees, and C. L. A. Clarke. Overview of the TREC 2013 Contextual Suggestion track. In *Proceedings of TREC'13*, 2013.

[6] A. Dean-Hall and J. Thomas Smarter Cities, Safer Travels: Integrating Contextual Suggestion In [2]

[7] R. Deveaud, M. Albakour, C. Macdonald and I. Ounis Challenges in Recommending Venues within Smart Cities. In [2]

[8] F. Kresin. Smart Cities, Smart Citizens and the case for the CitySDK. In [2]

[9] H. Kukka, V. Kostakos, T. Ojala, J. Ylipulli, T. Suopajrvi, M. Jurmu, and S. Hosio. This is not classified: everyday information seeking and encountering in smart urban spaces. *Personal and Ubiquitous Computing*, 17(1):15–27, 2013.

[10] Y. Li and A .Smeaton. From Smart Cities to Smart Neighborhoods: Detecting Local Events from Social Media. In [2]

[11] V. Lopez, S. Kotoulas, M. L. Sbodio, M. Stephenson, A. Gkoulalas-Divanis, and P. Mac Aonghusa. Queriocity: A linked data platform for urban information management. In *The Semantic Web–ISWC 2012*, pages 148–163. Springer, 2012.

[12] P. Mac Aonghusa. Content, Connection and Context: From Data to Insight in Smarter Cities. In [2].

[13] C. Manta-Caro and J. M. Fernández-Luna Modeling the Web of Things from an IR approach. In [2].

[14] D. Maxwell, S. Raue, L. Azzopardi, C. W. Johnson, and S. Oates. Crisees: Real-time monitoring of social media streams to support crisis management. In *Proceedings of ECIR'12*, 2012.

[15] A. Moralis, G. Perreas, A. Glaros and D. Dres. "Search-the-City" - A versatile dashboard for searching and displaying Environment and User Generated Content in the context of the future Smart City. In *Proceedings of ECIR'12*, 2012.

[16] D. Namiot On Mining Mobile Users Monitoring Logs. In [2].

[17] J. Oberlander. Tourists in Smart Cities: Data mining for hidden treasures In [2].

[18] R. Perego Mining Digital Footprints for Smart Tourism. In [2].

[19] Y. Ren, M. Tomko, K. Ong, Y. B. Bai and M .Sanderson. The Influence of Indoor Spatial Context on User Information Behaviours. In [2].