PersonisJ: Mobile, Client-Side User Modelling

Simon Gerber, Michael Fry, Judy Kay, Bob Kummerfeld,
Glen Pink, and Rainer Wasinger

School of Information Technologies, University of Sydney, Sydney, NSW 2006,
Australia
{sger6218,mike,judy,bob,gpin7031,wasinger}@it.usyd.edu.au

Abstract. The increasing trend towards powerful mobile phones opens many possibilities for valuable personalised services to be available on the phone. Client-side personalisation for these services has important benefits when connectivity to the cloud is restricted or unavailable. The user may also find it desirable when they prefer that their user model be kept only on their phone and under their own control, rather than under the control of the cloud-based service provider. This paper describes PersonisJ, a user modelling framework that can support client-side personalisation on the Android phone platform. We discuss the particular challenges in creating a user modelling framework for this platform. We have evaluated PersonisJ at two levels: we have created a demonstrator application that delivers a personalised museum tour based on client-side personalisation; we also report on evaluations of its scalability. Contributions of this paper are the description of the architecture, the implementation, and the evaluation of a user modelling framework for client-side personalisation on mobile phones.

1 Introduction

Personalisation has the potential to offer many benefits, particularly in reducing information overload by enabling a person to be more efficient in finding the information they need or want. Personalised systems can also be valuable in an active role, alerting the user to useful information. But there is a tension between such personalisation and privacy; the user model that drives personalisation is based upon the user's personal information. Moreover, there is evidence of considerable community concern about the proper protection of such information, for example [1].

One way to address such concerns is to perform the personalisation at the *client-side*, with the user's model stored on their own system. This is in contrast to the widespread *server-side* personalisation. Consider, for example, an e-commerce website such as 'www.amazon.com', where customers must register with the site in order to shop. The site can log every action they take while they are logged in, such as the items they view, add to their shopping carts and ultimately buy. This is used to create a user-profile which is held on the server. The website's owners are in control of this user model and the way it is used.

P. De Bra, A. Kobsa, and D. Chin (Eds.): UMAP 2010, LNCS 6075, pp. 111–122, 2010.
© Springer-Verlag Berlin Heidelberg 2010

In client-side personalisation, the user model is controlled by the user and the personalised applications that also run on their machine, under their control.

We now consider the issue of mobile personalisation. Mobile phones are providing an increasingly important interface for people to access information. There are currently over 3.5 billion mobile phone subscribers [2] and there is an increasing trend for these to have data subscriptions (for example, 40% of mobile users in Japan have a data plan). Interestingly, studies of how people are actually using their mobiles reveal that many use their phones for internet access, even while in their own homes or with another computer nearby [3].

At present, personalisation of the information delivered to mobile phones is typically performed at the server-side, by services in the *cloud*. This has been a necessity due to the limited computational power and memory of mobile phones. However, with widely-available consumer phones becoming increasingly powerful, it is becoming feasible to support client-side personalisation for these devices.

To support mobile personalisation on phones, we need new tools to support the creation of new applications. PersonisJ is one such tool, providing a framework for developing context-aware, personalised applications on a mobile phone. It can support reuse of user modelling information by arbitrary personalised applications running on the device. PersonisJ is unlike other personalisation shells or context-aware frameworks in that it treats the mobile device as a platform, rather than simply as an actor in a larger framework [4]. It also has to operate under very different constraints from previous user modelling frameworks, because it must take account of the power constraints for programs running on a mobile phone.

The next section reviews related work. We then describe the architecture and implementation of PersonisJ followed by our validation of it by demonstrating its use in the MuseumGuide application and our evaluations of its scalability. Finally, we discuss the implications of the work and future directions.

2 Related Work

The relatively recent emergence of powerful mobile phones has created new possibilities for mobile personalisation and the associated needs for personalisation. For example, studies point to the need for personalisation of mobile search interfaces [5]. There has been considerable exploration of mobile personalisation for a range of contexts and types of application. For example, personalisation of information available has been based on the user's social context [6] and location [7]. In e-commerce, personalisation has been widely deployed, with an increasing role for mobile, m-commerce [8]. Some interesting forms include systems that enable retailers to push recommendations to the mobile customer [9] and to offer both personalised product details and in-store customer advice [10]. Another important class of applications, mobile guides [11], can cover roles as diverse as museum guides, navigation systems and shopping assistants [12]. For example, the PEACH [13] system delivered personalised information about the art in a museum on PDAs.

Early research in mobile personalisation has been dominated by a view of the mobile phone as the client of (and portal to) a powerful server which was responsible for the personalisation, often restricted to a particular space such as a university or hospital [14]. In the mobile personalisation work described above, the architecture of the systems places personalisation at the server side. We have not found reports of mobile client-side personalisation, or even mobile applications that reuse frameworks for client-side personalisation.

One of the barriers for client side personalisation is the lack of a framework for the user modelling. A recent review of generic user modelling systems [4] points to the considerable work on such frameworks for server side personalisation.

PersonisJ is strongly influenced by the PersonisAD context-aware modelling framework [15]. Distinctive features of this modelling framework are: the same mechanisms model users as well as devices and places; it supports distributed user modelling, particularly important for pervasive computing applications; it provides *scrutable* modelling, meaning that it was designed, from its foundations, to support a user's scrutiny of their user model and the way that it is used. It is also able to perform lightweight user modelling, making it a promising foundation for the phone where power consumption is a major concern.

We now describe key elements from PersonisAD that are important for PersonisJ. PersonisAD represents a model as an hierarchy of *contexts* which can contain *components*. It is based on the *accretion/resolution* representation. Applications interact with PersonisAD via three primitive operations. The first looks up a Model for a particular person, device or place. The application can then can use a *tell* operation to supply evidence, and an *ask* operation to request the value of a component. This value is dynamically determined at the time of the *ask* based on a two part process. First, an *evidence filter* selects just the evidence allowed for the application which performed the *ask*. Then a *resolver* interprets the set of evidence. For example, a playlist application might use a resolver for a person's favourite genre from a list of evidence that includes the songs they have most recently played. Notably, flexibility and power in the reasoning comes from the availability of a range of evidence filters and resolvers. Some aspects of PersonisAD are not suitable for use on a mobile device. Notably, it is a distributed server application and *always on* so that clients can make TCP/IP connections. While the essence of PersonisAD gives a conceptual foundation for this work, the demands of creating a framework for client side personalisation on a mobile phone have meant that we created PersonisJ from scratch and independently from it. Unlike PersonisAD, written in Python, PersonisJ is written in Java and runs as a native application on Android.

3 Architecture and Implementation

We now describe the PersonisJ framework. We begin with the conceptual level, which has much in common with PersonisAD as described above. Then we present the high level architecture. The actual implementation was on the Open

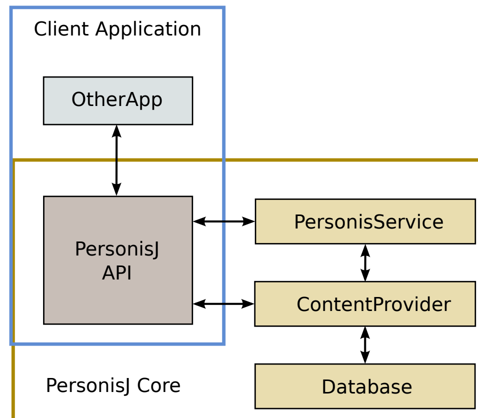

Fig. 1. PersonisJ Architecture

Handset Alliance's Android platform[1]. Pure implementation details are relegated to footnotes.

At a conceptual level, PersonisJ represents user models as an hierarchical structure of *contexts*, which can contain the *components* to be modelled. For example, it may have a context for the user's visits to museums and within this it may have components modelling the museums they prefer. Each component accretes evidence. This is essentially a value with metadata indicating how and when it was received. The metadata supports flexible evidence filtering and resolution and is a basis for supporting scrutability. The hierarchy of contexts and components constitute an *ontology* and a sub-tree within the hierarchy is a *partial ontology*.

Figure 1 shows the key modules of the PersonisJ architecture. The *PersonisJ API* enables an arbitrary application to access the user model, albeit only after the application has been granted read and/or write access to this model. *PersonisJ Core* provides the *API*, *Database*, *ContentProvider* and *PersonisService*. We now describe how each of these have been designed to represent the model and to secure access management.

The *Database*[2] has tables, with rows for each context, component and piece of evidence in the model. The *Database* is not accessible outside the *PersonisJ Core*. Instead, access is mediated by the *ContentProvider*, which specifies a unique "content URI" for content it exposes to the client application.

When a client interacts with the model, via the *PersonisJ API*, it must use a generic *ContentResolver* to act upon data, using content URIs. It is critical that client applications cannot directly modify anything within the PersonisJ database; for this reason, custom security permissions have been created for PersonisJ (READ_CONTENT, WRITE_CONTENT, and TELL_PERMISSION). These permissions are not intended to restrict unauthorised access; rather, it

[1] Android, `http://www.android.com`
[2] SQLite, on the Android platform.

forces applications to explicitly declare how they wish to interact with PersonisJ. All applications intending to access the PersonisJ database must state the permissions they require upfront[3]. To avoid direct third party application access to the database, the WRITE_CONTENT permission has been given the Android protection level of "signature", meaning that only applications signed with the same certificate as the PersonisJ code have direct write access to PersonisJ content. This has the effect that only two of the three above defined permissions are publicly available; third-party applications intending to read from the database use the READ_CONTENT permission (which corresponds to the *ask* operation) and those intending to contribute to the database do so via the PersonsisService module using the TELL_PERMISSION (corresponding to the *tell* operation).

To allow client applications to interact with PersonisJ, *PersonisJ Core* provides the *PersonisService* in Fig. 1. Because this resides within the same application as the *ContentProvider*, it can freely write to the PersonisJ *Database*.

The *PersonisService* is also in charge of handling imports. It is not possible to expose the ability to create new contexts and components via the *ContentProvider* without also exposing the ability to modify or delete them. This problem is solved by allowing client applications to pass the *PersonisService* a description of the partial ontology they require encoded in JavaScript Object Notation (JSON). The *PersonisService* then creates any necessary contexts and components.

PersonisService has one other responsibility. Any time it processes a *tell* operation it will broadcast a message[4] indicating which component was changed. It will then walk back up the context tree to the root, broadcasting notifications for each parent context in turn. This allows client applications to 'listen in', and discover if a particular component has received new evidence or, more generally, if any component beneath a particular context has changed.

The ontology is normally specified by applications. However, location is so fundamental that PersonisJ provides a predefined *location monitor* context in the phone model, with components for the co-ordinates of each location value. When turned on, this uses the phone's GPS to record any change in location as Evidence in the PersonisJ model.[5]

PersonisJ provides an Application Programming Interface [API] that applications can use to interact with the system. The *ask* operation returns a single value for a component. An important aspect of PersonisJ is that each component may have a list of evidence. In order to resolve multiple pieces of evidence into a single value the PersonisJ API provides a *Resolver* interface, which interprets a

[3] Security permissions are enforced by the Android OS, and the user is made aware of the required permissions when installing the application.

[4] An *Intent* in Android.

[5] The location monitor utilises the location API provided by Android. To conserve battery power, it does not turn on the GPS of its own accord. Instead, it hooks in to its operation whenever another program turns it on. The user can disable, or re-enable, the location monitor. While the GPS is disabled the location monitor does not reside in memory and uses no additional battery.

set of evidence, returning a single value. The API includes a selection of 'default' resolvers for numerical values, booleans, strings and dates. The actual resolution of values is executed within the client application, enabling them to provide their own *Resolver* implementations. The evidence passed to the *Resolver* can optionally be passed through an *EvidenceFilter*, which chooses the pieces of evidence to be used by the *Resolver*.

The API also exposes the ability to import and export partial ontologies. It provides a method, which can nominate any context, to gain the exported ontology. This operation occurs within the calling process. Optionally, a URL can be provided to the export function. After encoding into JSON, the generated string will be passed to the PersonisService which, in a background thread, will transmit the data to the specified URL via an HTTP POST. The import function works in a similar fashion, to upload the partial ontology defining a new part of the model with *PersonisService* performing this operation as external applications do not have write access to PersonisJ itself.

4 Evaluation

This section reports the two approaches we have used to validate the PersonisJ framework. First, we used it to create a personalised client-side application called MuseumGuide, which is able to notify a user of nearby museums and then download content for any museum that the user is keen to visit. Then we conducted scalability evaluations.

4.1 MuseumGuide Application

Consider the following scenario:

Alice and Bob, with their young family, are on a driving holiday in Sydney. Their phone has a model of the family's entertainment preferences including: low cost; suitable for children; kids are interested in ancient Egypt. The MuseumGuide, running on Alice's phone and aware of their location, sends an alert that they are near the Nicholson Museum. They decide to go to the museum and, on arrival, download a personalised museum tour based on a detailed model of the family's interests.

We now describe our implementation of MuseumGuide, making use of the PersonisJ framework to model a family's entertainment interests, as outlined in the scenario, and their more detailed interest model.

Figure 2 shows the MuseumGuide architecture, including the PersonisJ API module described earlier on (see Fig. 1), providing the application with access to the PersonisJ database. It operates as a client-side application, and like other Android applications requires the user to confirm when they are installing the application that they are happy with granting it the requested permissions. In this case, it requires access to the Internet, to the PersonisJ *ask* and *tell* operators, as well as several other permissions.

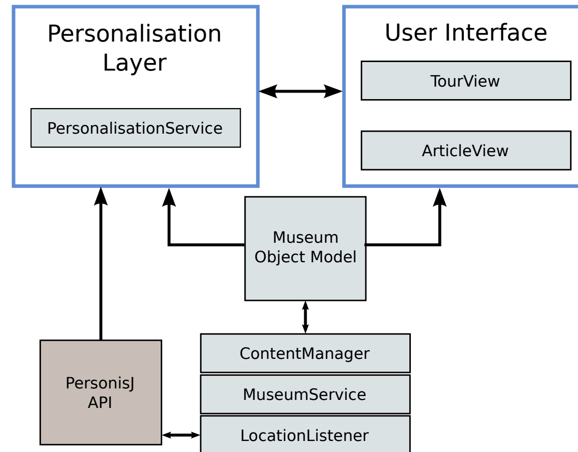

Fig. 2. MuseumGuide architecture

MuseumGuide uses a BroadcastReceiver, named `LocationListener`, to listen for component changes broadcast by PersonisJ. If the Location component changes, the alert is passed to `MuseumService`. This compares the new location against a stored list of museums. It sends a notification as shown on the left in Fig. 3. When viewed it starts an Activity that checks to see if there is content that can be downloaded (or updated) for the museum in question. If so, a prompt is displayed asking the user to confirm whether it should go ahead with the download (middle of Fig. 3).

The content is downloaded from a URL that is generated based upon the name of the museum. Once loaded, the background service invokes the `ContentManager` which imports it to MuseumGuide and then sends a second notification (right screen in Fig. 3). Responding to this notification brings MuseumGuide into the foreground.

At this point the content is ready for viewing. A `PersonalisationService` is situated, architecturally, between the ContentProvider and its UI. This service is responsible for personalising the content before it is displayed. Figure 4 shows one such personalised article. In this example, text content that has been adapted for a young child. We used content for the Nicholson Museum, located on the University of Sydney campus, taken from an existing museum guide [16].

The PersonalisationService exists as a separate Service to enforce a separation between the content, the user-interface and the personalisation of the content. The service takes a museum and article identifier as input and returns personalised output, based on a simple personalisation algorithm based on the age of the user.

The above description illustrated how PersonisJ enables applications to register for updates in a manner that is no more complicated than requesting an update from existing Android system services. Once the logic for resolving a value from evidence has been encapsulated within a Resolver class it requires only a single *ask* to retrieve it.

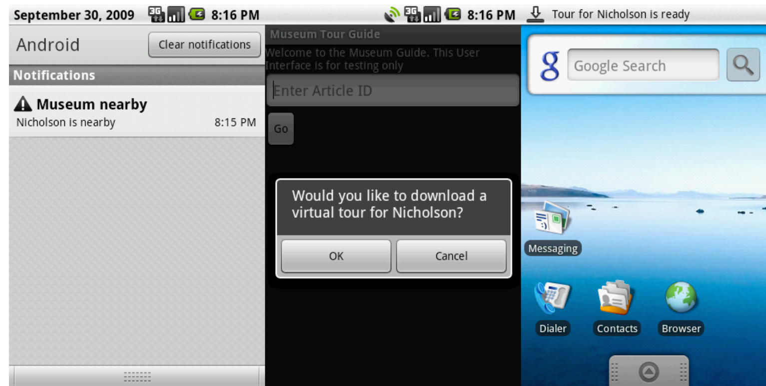

Fig. 3. MuseumGuide interface

In this way, PersonisJ makes it possible to create a context-aware application, which achieves its personalisation by accessing the user model via calls to a high-level API. We note that obtaining location updates from Android directly requires only slightly less code. However, PersonisJ provides a higher level of abstraction.

4.2 Scalability Evaluation

Before beginning our evaluation of PersonisJ we note that the DalvikVM[6] does not provide any 'Just in Time' [JIT] compilation. This has three important consequences. Firstly, it means managed code will always run slower on Android than its native equivalent. Secondly, it means that method level optimisations are important, as one cannot rely on minor inefficiencies being 'optimised away'. Finally, the lack of JIT compilation means we can be relatively naive about our performance testing, knowing that the code we write is, more or less, the code that Android executes.

Performance was measured using the profiler built into the Android framework, which has a resolution of micro seconds. The results are shown in Table 1. The critical column is the final column which shows the relative performance for the tested actions. This was calculated as the ratios of the average time per call normalised to one, then rounded to the nearest ten. The 'time' and 'average time' columns were included for completeness, but should not be relied upon as a measure of real-world performance.

Tests 1 to 4 involve actions selected from the sample actions in Table 1. This gives a baseline for PersonisJ performance against some simple operations. Tests 5, 6 and 7 reflect some basic PersonisJ API operations. Test 5 is the *tell* operation. We would expect it to be a relatively quick API call as it does not do much work itself, but rather encapsulates a call through to the PersonisService class in the core framework. Test 6 is the corresponding *ask* operation. Test 7 is

[6] The virtual machine used in Android.

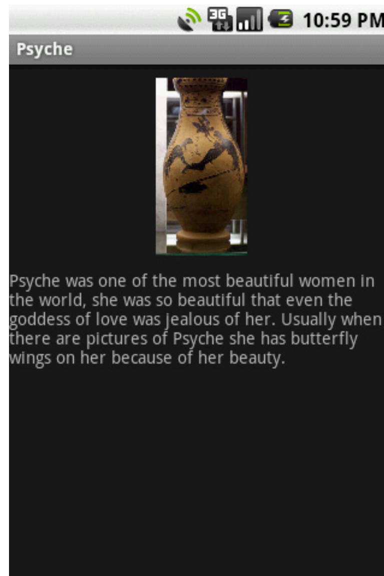

Fig. 4. Example of a personalised article

an example of an API call that would be used to drill down through an ontology. Finally, Test 8 retrieves the same information from the PersonisJ database as Test 7, but does so using a ContentProvider directly and is optimised to use only one query instead of two.

These results indicate that PersonisJ's fastest operation is clearly the *tell* operation. This is expected, as a *tell* operation is asynchronous. The *ask* operation, in contrast, is a lot heavier and can be expected to differ significantly based on the supplied parameters. The *ask* operation tested in Test 6 was as lightweight as possible. Only the latest piece of evidence was examined and a default String resolver was used. The String resolver performs no additional computation, as all evidence is stored internally as a String. Even so, this simple *ask* took approximately three times longer than a *tell*. Test 7 is an example of a fairly common class of method. The ability to retrieve a model, context and Component by name is part of the required setup any client application must perform before it can even start to call *ask* or *tell*. However, once acquired, these methods do not usually need to be called again until the next time the application is killed and then relaunched. Test 8 acquires the same data as Test 7, but does so using only a single query. Its performance is roughly half that of Test 7. From this, we can determine that when only a few columns are involved it is much more efficient to return all columns in the initial query. This translates into an easy optimisation that can be applied to all PersonisJ calls that require ContentResolver queries.

We can also see from the results that the PersonisJ operations are the same order of magnitude as inflating a trivial user interface, with just one button, from XML (Test 4). Any user interface of practical use would far take longer, We can see that an *ask* on the model (test 6) takes of the order of 0.1 seconds.

Table 1. Performance Tests: The result of 200 calls to various Android and PersonisJ methods, profiled using 'traceview'. Results are sorted by relative performance. An asterisk * before the test number indicates a test involving PersonisJ.

Test Description	#	Time (ms)	Time/Call	Relative Performance
String.length()	1	7.606	0.038	1
HashMapIterator.next()	2	70.822	0.354	10
HashMap.put()	3	142.474	0.712	20
PersonisContext.tell()	*5	5763.410	28.82	760
LayoutInflator.inflate()	4	13904.226	69.521	1830
'Raw' context query	*8	14719.000	73.595	1930
PersonisContext.ask()	*6	24472.835	122.356	3220
PersonisContext.getChildContext()	*7	28371.960	141.860	3730

We can therefore conclude that acquiring complex information from PersonisJ would have not add a noticeable delay to the loading of a typical interface screen.

The PersonisJ API takes up 44.7kb when converted to dex format. This further compresses to 20.414kb in a .jar file. The PersonisJ application is 37.24kb as an APK package. It takes up 108kb of space on the phone when uncompressed. These sizes are negligible when compared to most Android applications. The PersonisJ Core Framework has essentially no UI and no other packaged resources, such as images. This makes it smaller, when first installed, than all but the simplest of applications. However, PersonisJ does create a database which will grow continuously over time.

The data requirements of PersonisJ are not dictated by PersonisJ itself, but rather the client applications. A full analysis of disk space requirements is therefore neither possible nor particularly helpful. An empty PersonisJ database is 10KB (10240 bytes).

In summary, the time performance of PersonisJ is adequate for the tasks required of it. Simple operations and resolvers that do not require much evidence are fast enough to be performed in the main UI thread without affecting application performance. More complex resolvers and contextual reasoning should be performed in a background thread, but are unlikely to take longer to run than setting up a typical UI screen. With respect to space performance, the size of the PersonisJ API is modest. The PersonisJ database, however, could grow too large over a period of only a few months. Future versions of PersonisJ will address this, with options to move the database to other storage media, prune old evidence, or back up old data to secure, personal storage over a network.

5 Conclusions

The goal of PersonisJ was to provide a personalisation framework that could support *client-side* personalisation on a *mobile phone*. This was motivated by two important potential benefits. First, client-side personalisation enables the phone to deliver a personalised service even when the phone is not connected to

a data network. Second, it stores the user model on a device that is *controlled* by the user. We have shown the effectiveness of the PersonisJ framework in its capability to support the demonstrator application MuseumGuide. This made use of a model which illustrates some of the breadth of user modelling power of PersonisJ. The demonstrator used of the model for: location to determine the nearby museums; phone owner's *preferences* for inexpensive and educational entertainment which caused a recommendation for the free Nicolson museum; their *interest* in ancient Egypt which also affected recommendation of Nicolson; age of user of MuseumGuide, which affected presentation of museum information. We also showed that PersonisJ runs efficiently enough that a typical phone interface screen will load in essentially the same time, whether there is personalisation or not. We have shown that the space demands are modest for short term user models; however, for a long term user model, we still need to create mechanisms for archiving or removing parts of the model.

With the user model restricted to the mobile phone, the privacy of that model depends upon the effectiveness of the associated security model. As we have described in this paper, the PersonisJ architecture was carefully designed to address security issues. Notably, PersonisJ mediates all accesses to the model. (For the details of the implementation, see [17]).

Another element of security relates to the behaviour of the *applications* that a user loads onto their phone. This is outside the scope of this paper. However, it is clearly a critical issue. This is why we have conducted parallel work on a security framework [18] which enables the user to control what an arbitrary application is permitted to do. For example, the user can limit an application to have no communication outside the phone. Or it may simply restrict the application from exporting any information from the phone. This is essential if a user is to download an arbitrary application, such as a personalised museum tour, since it ensures that the application can provide personalisation, based on the user model on the phone, but cannot send any information outside the phone. Our MuseumGuide application operated within an environment controlled by the security environment.

Client-side personalisation provides an important foundation for life-long user modelling, in which the user is able to create, edit, reuse, and extend their user model throughout their digital life experiences. We have described PersonisJ, a user modelling framework that can support client-side personalisation on the Android phone platform. Contributions of this work are: the first architecture for a user modelling framework for client-side personalisation on mobile phones; and its validation in terms of a demonstrator application and scalability tests.

References

1. Shilton, K.: Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. Queue 7(7), 40–47 (2009)
2. Church, K., Smyth, B.: Understanding the intent behind mobile information needs. In: Proceedings of the 13th international conference on Intelligent user interfaces, Sanibel Island, Florida, USA, pp. 247–256. ACM, New York (2008)

3. Nylander, S., Lundquist, T., Brannstrom, A.: At home and with computer access: why and where people use cell phones to access the internet. In: Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA, pp. 1639–1642. ACM, New York (2009)
4. Kobsa, A.: Generic User Modeling Systems, 136–154 (2007)
5. Church, K., Smyth, B.: Who, what, where & when: a new approach to mobile search. In: Proceedings of the 13th international conference on Intelligent user interfaces, Gran Canaria, Spain, pp. 309–312. ACM, New York (2008)
6. Kjeldskov, J., Paay, J.: Public Pervasive Computing: Making the Invisible Visible. Computer 39(9), 60 (2006)
7. Bilandzic, M., Foth, M., De Luca, A.: Cityflocks: designing social navigation for urban mobile information systems. In: DIS '08: Proceedings of the 7th ACM conference on Designing interactive systems, pp. 174–183. ACM, New York (2008)
8. Goy, A., Ardissono, L., Petrone, G.: Personalization in e-commerce applications, 485–520 (2007)
9. Kurkovsky, S., Harihar, K.: Using ubiquitous computing in interactive mobile marketing. Personal Ubiquitous Comput. 10(4), 227–240 (2006)
10. Li, J., Ari, I., Jain, J., Karp, A.H., Dekhil, M.: Mobile in-store personalized services. In: ICWS '09: Proceedings of the 2009 IEEE International Conference on Web Services, Washington, DC, USA, pp. 727–734. IEEE Computer Society, Los Alamitos (2009)
11. Kruger, A., Baus, J., Heckmann, D., Kruppa, M., Wasinger, R.: Adaptive Mobile Guides, 521–549 (2007)
12. Krüger, A., Baus, J., Heckmann, D., Kruppa, M., Wasinger, R.: Adaptive Mobile Guides. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 521–549. Springer, Heidelberg (2007)
13. Stock, O., Zancanaro, M., Busetta, P., Callaway, C., Kruger, A., Kruppa, M., Kuflik, T., Not, E., Rocchi, C.: Adaptive, intelligent presentation of information for the museum visitor in PEACH. User Modeling and User-Adapted Interaction 17(3), 257–304 (2007)
14. Chen, G., Kotz, D.: A survey of Context-Aware mobile computing research. Technical report, Dartmouth College (2000)
15. Assad, M., Carmichael, D., Kay, J., Kummerfeld, B.: PersonisAD: distributed, active, scrutable model framework for Context-Aware services. In: LaMarca, A., Langheinrich, M., Truong, K.N. (eds.) Pervasive 2007. LNCS, vol. 4480, pp. 55–72. Springer, Heidelberg (2007)
16. Czarkowski, M.: A Scrutable Adaptive Hypertext. PhD, University of Sydney (March 2006)
17. Gerber, S.: PersonisJ: A Platform for Context-Aware, Client-Side, Mobile Personalisation. PhD thesis, Unversity of Sydney (2009)
18. Pink, G.A.: Safe Execution of Dynamically Loaded Code on Mobile Devices. PhD thesis, Unversity of Sydney (2009)