Meeting the needs of mobile users of information systems

David Mountain and Andrew MacFarlane

City University, London, UK.

Correspondence to: David Mountain, Dept Information Science, City University, Northampton Square, London, EC1V 0HB, UK. dmm@soi.city.ac.uk

Abstract

In this paper, the information needs of mobile individuals are contrasted with those of their static (desktop) counterparts. The information needs of users of a mobile information system, implemented in an outdoor recreational area, are first investigated in a two-part user needs study, which finds a strong geographic component to their information needs have a strong geographic component. Next, four geographic post-query filters which are described which attempt to meet these needs. These filters are spatial proximity (distance in space), temporal proximity (travel time), speed-heading prediction surfaces (likelihood of visiting locations) and visibility (locations that can be seen). Two of these filters (spatial proximity and speed-heading prediction surfaces) are implemented in a mobile information system are evaluated with users in an outdoor setting. The results of evaluation suggest that information that to which post-query geographical filters have been applied is more relevant than unfiltered information, and that users find information sorted by spatial proximity to be more relevant than that sorted by a prediction surface of likely future locations. The paper closes by suggesting that one of the contributory factors to the failure of location-based services to become widely adopted, could be a neglect of the distinct information needs of mobile individuals.

Keywords: information retrieval; geographic information retrieval; mobile computing; mobile information needs; geographic filters.

1. Introduction

The widespread acceptance and use of the Internet has been hailed variously as the "death of distance" [1] and the "death of geography" [2, 3] since it allows individuals to access vast reserves of globally distributed digital information, regardless of their proximity to individual sources. As a consequence, spatial constraints upon access to information associated with physical limitations on individual movement [4, 5] have become less important. As predicted by Openshaw and Goddard [6] nearly twenty years ago, you no longer have to travel to an information source, or wait for it to be sent to you, in order to access it.

However, the content of documents frequently refers to one or more locations within the physical world since, features tend to be found, and events occur, at specific locations [7]. Despite the wealth of *geographic context* that appears to be implicitly contained within documents [8], the current generation of Internet information retrieval engines - relying upon exact matching between query terms and the textual content of information sources - have no way of representing or comparing these *geographic footprints* [9]. More recently, a field of research known as geographic information retrieval (GIR) [10] has emerged to tackle the problems associated with the "spaceless Internet", by identifying geographic terms in documents and building geographic ontologies to identify the spatial properties of documents. When performing spatial queries of this kind, it is generally assumed that the user of the system will define the geographic footprint associated with their query explicitly in some way, either using text input (such as a city name, address or zip code [11]) or via map interaction [10].

In parallel with the Internet revolution, the development of consumer handheld devices, such as personal digital assistants (PDAs) and mobile phones, has led to a new computing paradigm, that of mobile computing [12]. Mobile computing constraints include limited screen real estate [13] and reduced interaction between user and device due to more constrained input mechanisms and the distractions of the outside world [14]. As a result, mobile computing use tends to be characterised by short, frequent, task-focused sessions [15], and these tasks are often of a fundamentally geographic nature such as routing [16], rendezvous [17], searching around one's location [18], proximity messaging between acquaintances (xxx cite), tracking of dependents, employees or resources (cite xxx), proximity advertising (cite xxx), location-based tariffs [19]: functionality that defines the emerging field of *location-based services* or LBS [20]

The importance of the geographic context of mobile users, combined with their reduced ability to interact with a device whilst on the move, makes a strong case for automated techniques to define the region that is relevant to an individual's query at a given time. Previous research in this area has generally considered this context to be invariant and a property of the information sought [21] or space itself [22] rather than the behaviour of the individual. A variety of *geographic filters*, defining the geographic footprint of a user's query, may be used as a post-query filter (xxx cite xxx). The application of these geographic filters offers the potential to reduce the volume of information delivered, and perform ranking based upon notions of *geographic relevance* [23].

1.1. Research Questions

The ambition to develop information systems that can satisfy the information needs of mobile individuals gives rise to a number of research questions which straddle various scientific disciplines, notably (Geographic) Information Science and mobile computing. The first concerns the information needs (xxx cite xxx) of mobile individuals, and how these contrast with the information needs of users of static desktop machines. Given that mobile individuals are, by definition, moving through space, the next question to consider is how best to represent the geographic footprint associated with a mobile individual's information query, and whether this representation is task or user dependent?

This paper will investigate these questions, building upon research conducted for the WebPark project, a three year collaboration of six European partners funded by the EU 5th framework [24], whose primary aim was to develop a mobile information system providing *geographically relevant* information to visitors to outdoor recreational areas. First, further background to geographic information retrieval will be given. Next, the methodology and main results of a study to designed to establish the mobile information needs of visitors to an outdoor recreational area will be presented. This will be followed by the description of a mobile, location-aware information system – developed using the WebPark architecture, and designed specifically to satisfy those needs, and the user evaluation of this system. The paper closes with a discussion and conclusions.

2. The geographic information retrieval problem

2.1. Information retrieval and relevance

Criteria for judging the relevance of information has been central to the discipline of Information Retrieval (IR) for nearly forty years [25]. The central aim of IR has been stated as the retrieval of information (such as textual documents or multimedia) relevant to some query that represents a person's information needs arising as a result of some particular task or problem at hand [26]. Traditionally the different IR algorithms have been evaluated primarily at a system level with little reference to the user [27], however relevance has been acknowledged to be;

- multidimensional, in that it can be measured on a number of different levels [26],
- dynamic, in that a user's information needs are mental constructs that vary through time [28], and,

- complex, but inherently measurable [29].

This discrepancy has led to criticism of the IR community for relying on traditional relevance criteria that are solely objective, considering only the relationship between retrieved documents and the query from a computing perspective, rather than considering subjective dimensions of relevance related to the person who's individual information needs led to the query being conducted [29].

A system of five distinct but interrelated manifestations or *levels* of relevance has been suggested [30] which aims to integrate the different frameworks for defining relevance that have emerged in different disciplines. *Algorithmic (1)* relevance considers the relationship between the query definition and retrieved information sources based upon a system's internal relevance criteria. It can be assessed by a user in terms of the effectiveness of different algorithms when inferring which documents were relevant. S*ubject (2)* relevance considers the topic defined by the query and the topic covered by information sources. It can be assessed in terms the relationship between the query definition and retrieved information sources and can be assessed in isolation of a user. In the algorithmic case, methods make some assumption about what is relevant and are measured by their effectiveness. For subject relevance, it is assumed that the subjects covered by both the query and information sources are known. *Cognitive (3)* relevance (or pertinence) considers the relationship of the user's state of knowledge and information needs allowing retrieved sources to be measured subjectively in terms such as informativeness, novelty and quality. *Situational (4)* relevance (or utility) considers the relationship between the user's context in terms of task or problem in hand and their situation, such as their activity or environment. It can be measured in terms of appropriateness (with respect to the user's problem), usefulness (in decision making) and reduction in uncertainty. The geographic component of relevance can be seen as a subset of situational relevance, filtering information according to their surrounding environment and their relationship to it – ie their geographic context. *Motivational (5)* relevance relates the user's goals and intentions to the retrieved sources and can be measured in terms of satisfaction and accomplishment [31].

2.2. Geographic Information retrieval

> ***"Geographic information is pervasive on the web"*** Silva *et al* (2004) [8]

Individuals performing natural language searches of information sources, for example using a web search engine, often require information that is geographically specific and will include ***toponyms*** (placenames) in the query to define its spatial focus [32]. The approach of the current generation of Internet search engines is exact matching of terms in the query and documents. When applied to geographic query terms, this approach can only retrieve exact matches but fails to consider documents containing nearby place names, alternative spellings and names, or placenames that refer to the same location at a different resolution [32]. Studies suggest that nearly 20 percent of Internet queries contain a geographic term [32], and it is to counter this lack of geographic intelligence that the research field of geographic information retrieval emerged.

In order to ensure that retrieved information is *geographically relevant*, the geographic extent of the query and individual records in the dataset must be known. The geographic extent of the query may be determined by including a toponym in a natural language search [33], by defining an extent on a map [34], or by determining an individual's current location using some positioning determining technology [35]. There are many approaches to determining the geographic extent of the candidate results of a query, which have developed to take account of the nature of the information repository that is to be searched. First, many Features of Interest (FOI) databases exist, designed to associate a geographic footprint with each feature in the database. Many of these come from the commercial field, where features – with particular emphasis on commercial services – are given a point reference, usually derived from address fields of a database, perhaps the most familiar being the Yellow Pages [36]. In the design of mobile information systems, georeferenced content is often created from scratch [37] or extracted from existing databases to create a library of spatially referenced multimedia documents [38]. For the majority of natural language documents that comprise the World Wide Web however, there is no one clear solution to assigning reliable spatial references to individual sources, and many approaches to extracting geographic information from unstructured text have been suggested [39].

Geographic information is found throughout Internet resources. A study by Silva et al. [8] suggests that web documents contain an average 2.17 references to geographic entities. The first stage of geographic information retrieval is to recognise these geographic references or toponyms within the free text – known as *geoparsing*. Next, a geographic footprint must be associated with the toponym – a process known as geocoding [40], which may be achieved by using a gazetteer [39]. Many complications exist, such as the case where multiple toponyms exist in a single document, or where the same toponym refers to multiple locations (for example "London, UK" and "London, Otario"), and much current research in this area is concerned with the *disambiguation* of these terms to identify the true geographic focus of free text documents [21].

Some researchers have constructed *geographic ontologies* which provide a model of geographic terminology and structure of space [41]. These ontologies act as geographic knowledge bases, defining the relationships between toponyms, such as whether two locations are adjacent, or one is contained within the other [8]. This has been done not only for formal placenames with known administrative boundaries, but also imprecise regions with no formally defined spatial extent such as "The Mid-West" or "The West Midlands" [42]. In this way, the geographic extent of a document can be expanded in beyond the topynyms found in the text, to include surrounding and containing regions. Once the geographic extent of both query and records in the dataset has been defined, the two can be compared to give spatial similarity score that allows records to be ranked according to their relevance to the query [43]. In the simplest case this may be a Boolean classification, according to whether the extents of query and record overlap or not. Various authors have attempted definitions of spatial similarity, which is usually based upon the degree of overlap of two extents and assumes that the spatial extent of both query and record can be precisely defined [43].

A little investigated area with respect to document ranking is one which examines the evidence not only from spatial data, but also from the users underlying information need. There are many term weighting models for information retrieval which utilise textual relevance, such as BM25 in the Robertson/Sparck Jones Probabilistic model [ref]. Van Kreveld et al [ref] consider the both spatial and textual relevance using multi-dimensional scatter ranking methods in order to combine evidence from both sources, rather than produce a single document score for a rank. The issue of whether to use these ranking fusion techniques or single score is a open question. In the former the problem is being able to combine the relevance information such that no one source dominates the other, whereas in the latter we need to manipulate variable statistics (e.g. term frequency) which has significant theoretical and practical implications. We believe this is a significant research area.

References:

Van Kreveld, M., Reinbacher, I., Arampatzis, A and Van Zwol, R. (2005). Multi-Dimensional Scattered Ranking methods for geographical information retrieval, GeoInformatica 9(1), 61-84.

Robertson, S.E. and Sparck Jones, K. (1976). Relevance weighting of search terms. Journal of the American Society for Information Science, 27, 129-146.

2.3 Mobile information retrieval

The convergence of computing and mobile telephony has led to more sophisticated handheld computers that have fast processors, larger colour displays, wireless network access, and increasingly, location-aware capabilities, utilising terrestrial and satellite-based position determining technologies [44]. This enabling technology is promoting the growth of context aware computing [45] and information retrieval systems that are able to take account of the location and behaviour of the user and apply this as a filter to improve the *situational relevance* of the information received by the user.

Several systems have already been developed that are able to filter information according to the user's location. A variety of technologies have been employed such as dedicated wireless hotspots at specified locations, serving information to about local features to suitably equipped mobile devices in the area [35]. Other systems have attempted a more generic solution where location-aware mobile devices access bespoke georeferenced datasets via a wireless connection and filter information according to the user's current context [46]. No systems thus far have succeeded in using the user's geographic context to permit the retrieval of geographically relevant information from the unstructured web sources.

3. Mobile Information Needs

Early in the design phase of the WebPark project [24], an extensive two-stage user needs study was conducted within the Swiss National Park, a project partner and a test bed for development and implementation. First, a questionnaire survey (n=1597) was conducted to establish visitors' primary sources of information once in the park. Respondents identified information boards (included in 66% of responses) as the primary source of information during visits to the park. This was followed by guide books (51%), nature trails (41%) and personal contact with staff (40%). Information boards, an inherently location-based approach, are strategically placed to offer information about the surrounding area. The dependence of respondents upon these boards, suggests that there is a strong geographic component to visitors' information needs. Similarly, following nature trails, being in contact with staff and attending guided tours all have a strong geographic component, usually providing information related, either directly or indirectly, to the surrounding area.

Table xxx. Information provision preferences when visiting the Swiss National Park.

	Rank	% responses
Information boards	1	66
Literature	2	51
Nature trail	3	41
Personal contact with staff	4	40
No info	5	19
Personal experience of SNP	6	19

Guided tour	7	14

Based upon 1579 valid responses

The second stage of the user needs study was a series of visitor shadowing exercises, in which visitors were accompanied during their visit to the park by a member of the WebPark project team who recorded all their questions over the duration their visit. The intention was to act as a "human mobile information system", recording and responding to queries, but never prompting the visitor for information, in an attempt to assess mobile information needs and anticipate potential scenarios of use [47]. During the user shadowing exercises, a total of 90 questions were recorded, of which 53 had some spatial component. Of the 53 questions with a spatial component, 15 were related to navigation (eg "Where are we?", "Where is the trail leading to Munt La Schera?") or landscape (eg "Is this Lake Livigno?"). Of the remaining questions with a spatial reference, 9 were related to park fauna (eg "Are there any marmots here?"), 12 to flora (eg "At what elevation is the timberline? Will we pass it on our way?"), and 8 to the geomorphology (eg "Is this moraine?") [47]. In many questions, the spatial component of the query was implicit as the question referred to the current location (eg "... around here … "), a visible location (eg "Is this… ?", "Is that …?") or a future location on the route (eg "Will we pass …?").

This implication that roughly 60% of queries have a spatial component is an interesting result that is at odds with studies of desktop users found in the literature, which suggests that only about 20% of queries have a spatial component [32, 48]. The key differences between this particular user needs study and those in literature contrast the information needs of mobile and static (desktop) users: this study was conducted in an outdoor environment; the participants were mobile; and user needs were recorded from comments and questions made by the participant, rather formulating queries that were submitted to, and logged by, an Internet search engine. The limited size of this user needs study means that care must be taken in interpreting results, however the discrepancy of this study with studies on desktop users could be that mobile individuals' information needs are more likely to be a product of their surroundings, and the environment in which they are interacting. An alternative interpretation is that static users of Internet search engines are *as* likely to have a spatial component to their information needs, but they do not include a toponym or spatial relation in the query having learned from experience that this spatial component is poorly handled by search engines [32].

It was clear from this study that the information needs of mobile individuals frequently contained a geographic component, that many queries would be meaningless without this geographic context, and that the majority of such queries could not be satisfied by existing information systems. The results of this study led to contemplation about the geographic filters that could be applied to increase the relevance of retrieved information [47].

4. Implementation

4.1. The WebPark mobile information system

Xxx Brief description of content, search capabilities, content, architecture, and screen grabs xxx

4.2. Geographic filters for mobile information retrieval

The mobile information needs section closed by proposing geographic filters as an approach to increasing the overall relevance of the information retrieved by users of mobile information systems. The filters aim to represent the geographic footprint associated with a user's query, and can be represented as continuous surfaces, or geographic features that bound the region of space considered relevant for an individual at some time, based upon some quantifiable geographic criterion. They can be applied post-query to restrict results to those which are not only judged to be relevant at the subject level, but also at the situational level - specifically in terms of geographic content.

Various assumptions are made about how information can be processed to be made more relevant to the geographic context of the user, based upon the geographic component of the mobile information needs revealed during the user shadowing exercise. The first is *spatial proximity*, where the closer an information source is to an individual, the more relevant it is. The concept of spatial proximity can be represented with a (spatial) buffer around an individual's location, or a surface that can differentiate based upon the degree of separation from that individual. based upon distance from the user's position or some other location. These filters are familiar from "where's my nearest" directory services [20] such as Yell [49] or Vodafone's Find and Seek [50]. A more sophisticated assumption is that of *temporal proximity*: regions that can be reached in a shorter period of time are more relevant than those that are temporally distant. Isochrones can be used to represent the region of space accessible within a given time, and these travel time estimates may be calculated from analysis of the transportation network, or from geographic data mining of the previous spatial behaviour exhibited by users of the mobile information system [51]. Such an approach allows a query to specify a distance filter measured in minutes rather than metres. The next assumption is that information with a geographic footprint that coincides with an individual's likely future locations is more relevant than that located in places they are unlikely to visit. Spatial and temporal proximity can provide an effective prediction of future locations, in the absence of further information, however *speed-heading prediction surfaces* [52] based upon personal, recently exhibited spatial behaviour can take account of the general direction that someone is travelling in, the speed of travel, and the degree of sinuosity in their path [52]. The final assumption is that information that is visible is more relevant than that which is concealed. This geographic component was again revealed from analysis of the geographic component of queried made by visitors to the Swiss National Park, an alpine region with extreme relief where ridges and peaks can conceal many features that are spatially close. However the manmade environment has similar characteristics where the visibility is often severely restricted by the “urban canyons” of inner city roads, and highly visible and distinctive tall buildings are used for as navigation aids [53] in much the same way as hikers use distinctive mountain peaks.

For testing with end-users, each filter was given a more popular name that was intended to be more descriptive for this intended audience: the spatial proximity filter was given the popular name “search around me; the temporal proximity filter was renamed “accessible places”; the speed-heading prediction filter was renamed “search ahead” and the visibility filter renamed “visible places”. Each filter is described in table xxx and shown in figure xxx.

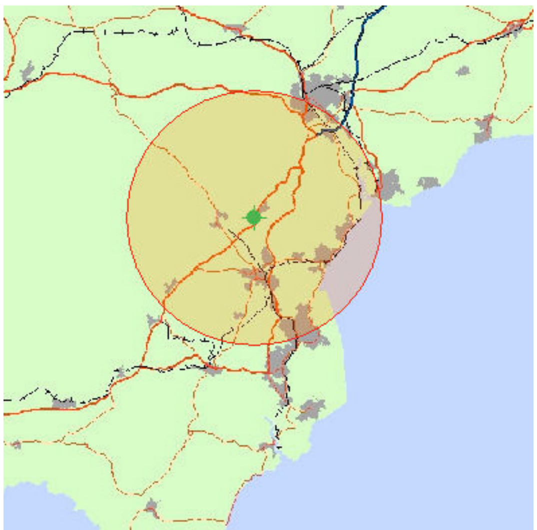
Table xxx. Geographic filters for mobile information retrieval

Name	Filter criteria	Potential query scenario
"Search around me"	Spatial proximity: Euclidean distance	"What animals are found around here?"
"Accessible places"	Temporal proximity: Travel time isochrones	"Which bus stops can I reach within 30 minutes?"
"Search ahead"	Speed-heading prediction surface: Likelihood of future path coinciding with information sources	"Which flowering plants am I likely to see in the next 30 minutes?"
"Visible places"	Viewshed analysis	"What is the name of that mountain peak?"

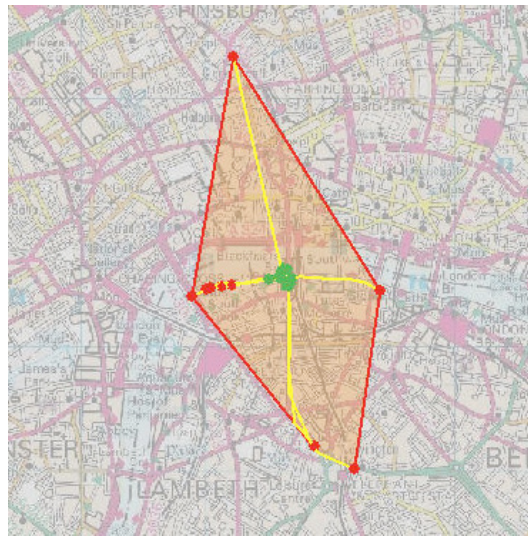
In the WebPark mobile information system as well as specifying the semantic component to a query, by selecting categories from a drop down list or performing a free-text search, the search could also be restricted by geographic criteria. Only two of the four post-query geographic filters described above were implemented. The "search around me" filter allowed users to restrict the extent of their search according to a specified distance from their current location (see figure xxx). The "search ahead" filter allowed users to restrict results to those which were they likely to pass within a given time limit. Users could also choose to "search in all park", which placed no geographic filter upon results. Results were ranked by distance from the user's current position in the case of the "search around me" filter, and by likelihood of coinciding with an information source, based upon the personalised prediction surface, in the case of the "search ahead" filter.

Figure xxx. Implementations of geographic filters

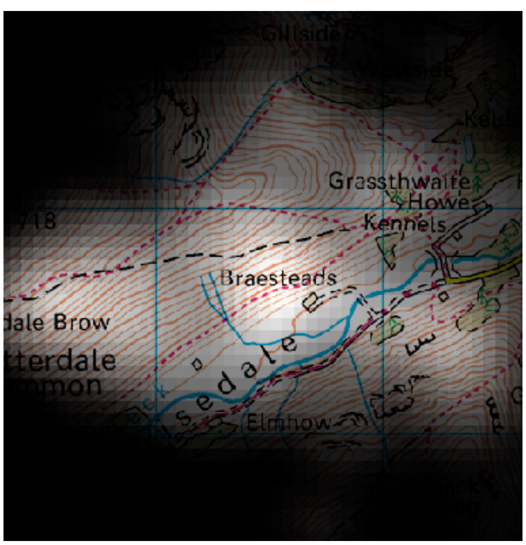
a. Search around me: Spatial proximity - 15km buffer (red line) around present location (green cross)



b. Accessible places: Temporal proximity - locations that can be reached within 10 minutes travel time



c. Search ahead: locations likely to be visited in next 30 minutes (light areas – high probability)



d. Visible places: visible locations (shown in dark red) from present position (green cross)

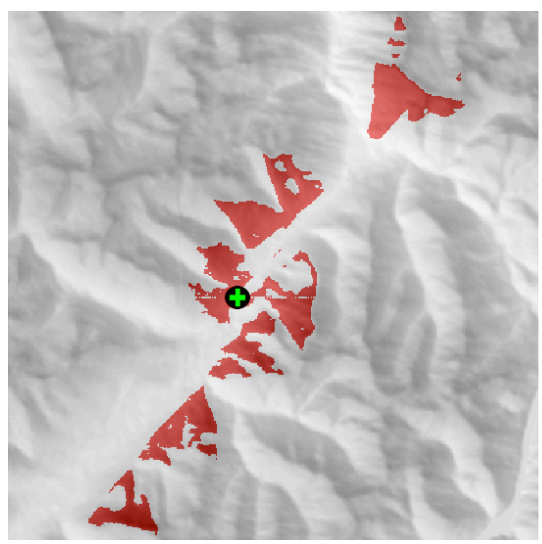
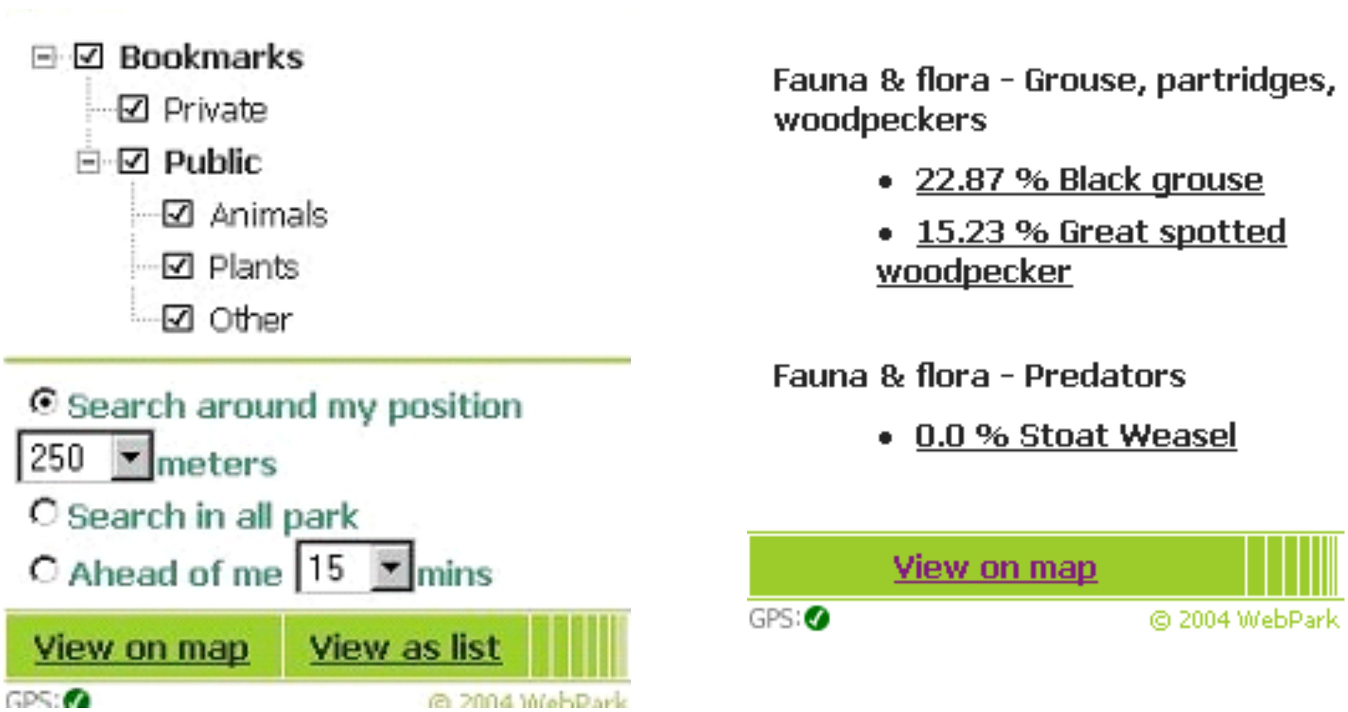


Figure XXXX: The WebPark search interface

a. The search interface allowing a user to specify both semantic and geographic components to a query

b. Results ranked according to the likelihood of an individual passing the geographic footprint associated with features of interest



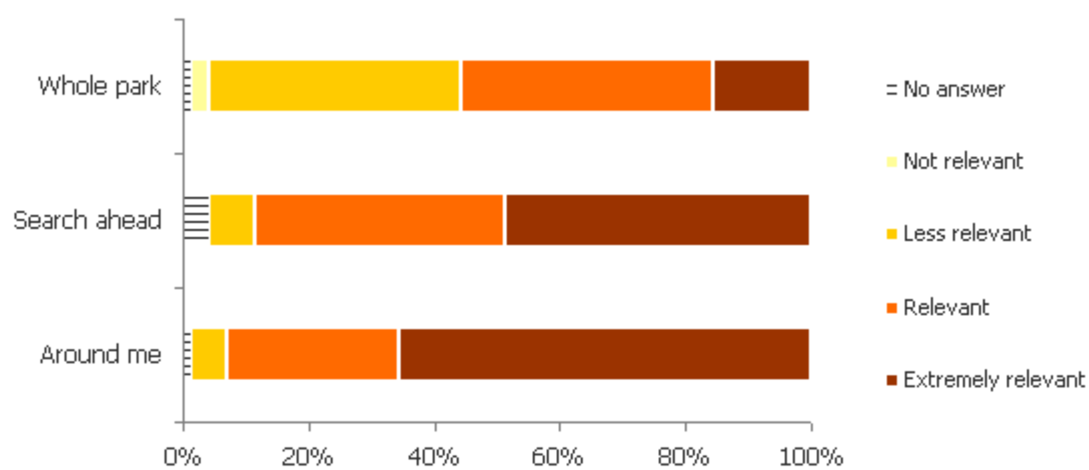
5. Evaluation

At the end of the WebPark project, a formal period of user evaluation was conducted using the Swiss National Park as a testbed. All project members had an input into the user testing strategy, and all attended the week-long user test session in the Summer of 2004. Test participants were given a brief introduction to the system at the National Park Information Centre, then left to use the device unaccompanied in the field for the duration of the day. Upon return, they were asked to complete a feedback questionnaire - a total of 87 questionnaires were completed [54]. The main advantage of this approach was the realism of the testing scenario, since the system was used *in situ*, as required, and without any intervention from the project team. The main disadvantage was that it could not be guaranteed that all users would utilise all of the system functionality, hence the response rate to some questions was low. Only questions directly relevant to this research, assessing the performance of alternative geographic filters for mobile information retrieval, are considered here.

The feedback from users about the geographic filters can be seen in Figure xxx. The "search around me" filter was considered to provide the most relevant information with two thirds of the respondents saying that this provided "extremely relevant" results, and over 90% claiming that results were either "extremely relevant" or "relevant". The "search ahead of me" filter also performs well, with a similar number (89%) claiming that results were either "extremely relevant" or "relevant", and half claiming that the results were "extremely relevant". The geographic filters clearly outperform the "search whole park" option, where 43% of respondents claimed that the information provided using this option was either "less relevant", or "not relevant" [54]. Considering the geographic component of the user questions identified in the visitor shadowing study it is unsurprising that the spatial proximity filter was considered to provide the most relevant results, since the majority of the questions referred to the user's current location for geographic context. The good performance of the "search ahead" filter suggests that this can also perform well, and that people are receptive to the idea of alternative notions of geographic context when making queries from mobile devices.

Figure xxx: Relevance of results using different geographic filters

Responses to the question "How relevant did you find the results using the following search functions?", from WebPark Summer testing 2004, in the Swiss National Park. The questionnaire was completed by 87 respondents.



Visitors were also asked about further geographic filters which were not implemented in the WebPark system (see Figure). These were;

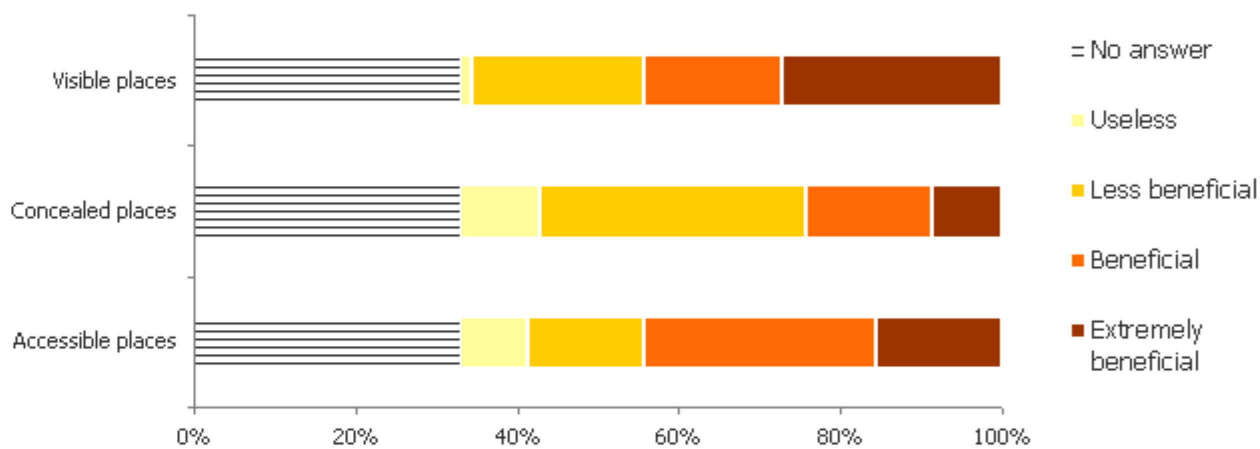
A “visible places” filter: retrieving only results which in the user’s line of sight;

A “concealed places” filter: retrieving results which are close by, but concealed from view by the terrain;

An “accessible places” filter: ranking results by the time taken to travel to the associated location.

Figure xxx: Perceived benefit of alternative geographic filters

Responses to the question “How beneficial would you find the following search options?” from WebPark Summer testing 2004, in the Swiss National Park. The questionnaire was completed by 87 respondents.



Responses suggest that the visibility filter would provide the greatest benefit to mobile users, with one quarter of respondents finding this “extremely beneficial”, and 44% considering it “extremely beneficial” or “beneficial”. A travel time (temporal proximity) filter is also considered to be of use, with a similar number (45%) considering this “extremely beneficial” or “beneficial”, although fewer (16%) chose the top category. The concealed option is considered to be of the least benefit with only 25% considering this “extremely beneficial” or “beneficial” and one third thinking it would be “less beneficial” [54]. Interestingly one third of respondents did not answer this question, suggesting that the use of these geographic filters may still be different to imagine, and hence difficult to express an opinion about.

6. Discussion and Conclusions

Within the field of mobile telecommunications, the focus for information retrieval on mobile devices has thus far been the emerging field of location-based services (LBS). LBS are, by definition, concerned with providing information about the services that are in some way geographically relevant to your current location: an implication of this approach is that you wish to purchase some service, be it consumables from a nearby store, or geographically relevant information from a service provider. In a taxonomy of mobile location services, Giaglis [55] identified six broad categories of services: emergency, navigation, information, advertising, tracking and billing. The commercial aspect of the majority of these services is clear, even in the need to commodify information services, where the subclasses “travel services”, “mobile yellow pages” and “infotainment services” were identified. There has been far less attention to searching for freely available information sources using the mobile Internet, or attempts to compare the similarity of the geographic footprint of free text web documents with the geographic footprint associated with a user’s query. This is at odds with the experience of the “static” Internet, where search engines rank web documents according to the relevance to the user query: the emphasis here is on providing relevant *information* (which may contain information about services), hence people are accustomed to accessing information sources without the barrier of cost. In the world of mobile information retrieval, so far services can search a subset of information – databases of services. None have successfully tackled the larger problem of retrieving geographically relevant information from free text documents found on the web.

There are several possible reasons why mobile information retrieval has thus far been primarily concerned with services rather than information sources. First, it has proved easier to identify the geographic footprint of many services than information sources generally: so called “service-scopes”. As discussed in the literature review, many databases exist detailing available services, with spatial information linking that service to an unambiguous location in the physical world [49, 56, 57]. All require a significant investment in terms of initial creation and subsequent maintenance and curation to ensure the currency and accuracy of the information listed. The notion of geographic relevance in mobile information systems that use these features of interest databases is usually based upon spatial proximity – which could be Euclidean or network distance - between a point location defining the user query (for example a toponym or postcode), and the point location representing the location of the required services [41].

Geographic information is also found in web resources: as mentioned in the literature review section 2 [8]. The research conducted within the field of geographic information retrieval suggests that the geographic footprints associated with web resources are far more ambiguous and complex than those associated with services in point of interest databases. One of the main reason cited for the complexity of the geographic footprints associated with web resources is that documents will frequently refer to more than one toponym in the text [21]. Where these multiple toponyms display clustering in space, this may assist in identifying the geographic focus of these resources, however when they are geographically distant, the presence of multiple toponyms becomes problematic. A further complicating factor, that has received less attention amongst the geographic information retrieval community, is that whilst gazetteers and geocoding tools can assist us in identifying the locations to which toponyms and addresses refer to in spatial coordinates, many natural features are not stored in gazetteers and hence cannot processed in the same way. The process of identifying the geographic footprint of natural feature is more complex, since many geographic phenomena, such as mountains and forests, do not have discrete boundaries but can display the characteristics of continuous phenomena on the one hand, or there may be no true defensible consensus as to what constitutes certain geographic concepts on the other [61, 62]. All these factors have made the identification of the geographic footprint associated with a web document very problematic.

A further reason for the focus of services in mobile information retrieval is likely to be related to the more commercial nature of mobile telecommunications when compared to the fixed Internet. The physical infrastructure of the Internet is not owned by any one single organisation, but is a network of networks, run collaboratively by a large number of public and private organisations [63]. The majority of mobile telecommunications networks, however, are designed to run independently and are owned privately by profit making organisations [64-68]. These organisations have made significant investment in both the network infrastructure, and the license to build third generation networks [69]. Given this context it is perhaps unsurprising that the mobile Internet, accessed via the networks of privately owned organisations, has concentrated on services that can be purchased at a price, as opposed to information that can be accessed for free.

Nevertheless, this promotion of location-based services has failed to see the take-up anticipated when they were first proclaimed as the “killer app” of the mobile Internet [70]. Various reasons have been suggested as to why these services have failed to perform as expected, including slow and imprecise positioning technology, and lack of processing power and functionality in client handsets, however the primary reason for the slow take up identified in a market report by Berg Insight AB [70] was:

“… the services offered up until today have simply been too slow and complicated to use.” [70]

Unlike desktop users, who can devote their full attention to a task with limited external distraction, mobile usage tends to take place in an environment where the cognitive load on the user from external sources is higher: rather than a blank wall behind their screen there is a dynamic world, within which they are moving and interacting.

Research suggests distinct difference in usage behaviour between desktop and handheld computing, with handheld usage being characterised by multiple, short sessions, as opposed, to the few, long sessions displayed in desktop usage [15]. This suggests that mobile users may be more in need of filters, to ensure the relevance of the information that they receive, than their static counterparts, since they have neither the time, nor the same level of attention, to manually refine information searches themselves, in the way that users accessing the Internet over static desktop devices have become accustomed to. A variety of geographic filters, deployed in different situations, may assist in increasing the relevance of retrieved information, and may possibly lead to the anticipated take-up of mobile information retrieval and mobile location services.

In order to retrieve geographically relevant information, relevance ranking based upon the similarity of the geographic footprints of Internet resources (or services), and the geographic footprint associated with a user's query must occur. Users of static machines, with mouse interfaces and keyboard interfaces may define the spatial extent of their query in a number of ways., for example clicking or dragging points, lines or areas on a map [10], or typing a toponym or postcode [49]. Whilst these options are also available to mobile users, there may be more automated and intuitive ways of identifying the geographic footprint associated with their query.

Acknowledgements

Much of this research was inspired by the WebPark project, which for three years offered an incredibly rewarding environment in which to work, and the xxx author(s) xxx would like to express gratitude to both to the EU IST 5th framework who funded the project and the individuals involved. In particular, the Swiss National Park made an great contribution in terms of assessing user needs, and conducting the user evaluation of the WebPark system.

References

[1] F. Cairncross, The Death of Distance: How the Communication Revolution Will Change Our Lives, *American Quarterly* 51(1) (1999) 160.

[2] S. Bates, The Death of Geography, the Rise of Anonymity and the Internet, *International Information Communication and Education* 18(2) (1999) 261-265.

[3] S. Bates, The Death of Geography, The Rise of Anonymity and the Internet, *International Information Communication and Education* 19(1) (2000) 122-126.

[4] M.P. Kwan, Gender differences in space-time constraints, *Area- Institute of British Geographers* 32(2) (2000) 145-156.

[5] B. Lenntorp, A time-geographic simulation model of individual activity programmes. In: T. Carlstein, D. Parkes and N. Thrift (eds), *Timing Space and Spacing Time Volume 2: Human Activity and Time Geography,* (Edward Arnold, London, 1978).

[6] S. Openshaw and J. Goddard, Some implications of the commodification of information and the emerging information economy for applied geographical analysis in the United Kingdom, *Environment and Planning A* 19((1987) 1423- 1439.

[7] P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind, Systems, Science, and Study. In: D. W. Rhind (eds), *Geographical information systems and science,* (John Wiley and Sons, Chichester, 2001).

[8] M.J. Silva, B. Martins, M. Chaves and N. Cardoso, Adding geographic scope to web resources. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[9] M.F. Goodchild, Towards a geography of geographic information in a digital world, *Computers Environment and Urban Systems* 21(6) (1997) 377-392.

[10] C.B. Jones, R. Purves, A. Ruas, M. Sanderson, M. Sester, M. van Kreveld and R. Weibel, Spatial information retrieval and geographical ontologies. An overview of the SPIRIT project. In: *SIGIR'02* (ACM, Tampere, Finland, 2002).

[11] Google, *Google local* (2004). Available at http://www.google.co.uk/lochp (accessed 29 Apr 2004).

[12] A. Helal, B. Haskell, J.L. Carter, R. Brice, D. Woelk and M. Rusinkiewicz (eds), Any Time, Anywhere Computing: Mobile Computing Concepts and Technology (1999).

[13] S. Brewster, Overcoming the Lack of Screen Space on Mobile Computers, *Personal and Ubiquitous Computing* 6(3) (2002) 188-205.

[14] L. Passani, Building Usable Wireless Applications for Mobile Phones, *Lecture Notes in Computer Science* 2411) (2002) 9-20.

[15] J. Ostrem, *Palm OS user interface guidelines* (2002). Available at http://www.palmos.com/dev/support/docs/ui/uiguide_front.html (accessed 29 Jan 2003).

[16] M. Kulju and E. Kaasinen, Route guidance using a 3D city model on a mobile device. In: *Mobile HCI02 Workshop: Mobile Tourism Support Systems* (Pisa, Italy, 2002).

[17] D. Chincholle, M. Goldstein, M. Nyberg and M. Eriksson, Lost or Found? A Usability Evaluation of a Mobile Navigation and Location-Based Service, *Lecture Notes in Computer Science* 2411) (2002) 211-224.

[18] J. Kjeldskov, "Just-in-Place" Information for Mobile Device Interfaces, *Lecture Notes in Computer Science* 2411) (2002) 271-275.

[19] A. Brimicombe and Y. Li, Dynamic Space-Time Envelopes for Location-Based Services. In: *Visualization and multimedia: Wireless GIS and Location Based Services* (University of Bristol, Harbourside, Bristol, 2004).

[20] Open Geospatial Consortium, *Open Location Services - OpenLS* (Open Geospatial Consortium, 2005).

[21] E. Amitay, N. Har-El, R. Sivian and A. Soffer, Finding the geographic focus of web-pages. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[22] R. Jose, A. Moreira, H. Rodrigues and N. Davies, The AROUND Architecture for Dynamic Location-Based Services, *Mobile Networks and Applications* 8(4) (2003) 377-387.

[23] J.F. Raper, Geographic Relevance. In: *Meeting on Fundamental Questions in GIScience* (Palace Hotel, Manchester, 2001).

[24] WebPark, *Geographically relevant information for mobile users in protected areas* (2005). Available at http://www.webparkservices.info (accessed 27 April 2005).

[25] C.W. Cleverdon and E.M. Keen, *Factors determining the performance of indexing systems* (Aslib Cranfield Research Project, Cranfield, 1966).

[26] A. Spink, H. Greisdorf and J. Bateman, From highly relevant to not relevant: examining different regions of relevance, *Information Processing and Management* 34(5) (1998) 599-621.

[27] P. Vakkari and N. Hakala, Changes in relevance criteria and problem stages in task performance, *Journal of Documentation* 56(5) (2000) 540-562.

[28] P. Ingwersen, Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory, *Journal of Documentation* 52(1) (1996) 3-50.

[29] P. Borlund and P. Ingwersen, The development of a method for the evaluation of interactive information retrieval systems, *Journal of Documentation* 53(3) (1997) 225-250.

[30] T. Saracevic, Relevance reconsidered. In: *Information science: Integration in perspectives. Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2)* (Copenhagen (Denmark), 1996).

[31] T. Saracevic, Interactive models in information retrieval (IR): A review and proposal., *Proceedings of the 59th Annual Meeting of the American Society for Information Science* 33((1996) 3-9.

[32] M. Sanderson and J. Kohler, Analyzing geographic queries. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[33] M. Nissim, C. Matheson and J. Reid, Recognising geographic entities in Scottish historical documents. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[34] F.C. Gey and K. Carl, Geotemporal Access to Multilingual Documents. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[35] A. Goker, H. Cumming and H.I. Myrthaug, Content retrieval and mobile users: An outdoor investigation of an ambient travel guide. In: *Workshop on Mobile and Ubiquitous Information Access, Mobile HCI 04* (University of Strathclyde, University of Strathclyde, Glasgow, Scotland, 2004).

[36] T. Sagara and M. Kitsuregawa, Yellow page driven method of collecting and scoring spatial web documents. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[37] N. Davies, K. Cheverst, K. Mitchell and A. Friday, Caches in the Air: Disseminating Tourist Information in the GUIDE System. In: *Mobile computing systems and applications: WMCSA '99* (Ieee, New Orleans; LA, 1999).

[38] A. Edwardes, D. Burghardt and R. Weibel, WebPark - location based services for species search in recreation areas. In: *21st International Cartographic Conference & 12th General Assembly of the INTERNATIONAL CARTOGRAPHIC ASSOCIATION* (Durban, SA, 2003).

[39] L. Hill, M.F. Goodchild and G. Janee, Research directions within the georeferenced IR based on the Alexandria Digital Library Project. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[40] R. Purves, P. Clough and H. Joho, Identifying imprecise regions for geographic information retrieval using the web. In: *GIS Research UK Conference* (University of Glasgow, University of Glasgow, Glasgow, 2005).

[41] C.B. Jones, A.I. Abdelmoty, D. Finch, G. Fu and S. Vaid, The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing, *Lecture Notes in Computer Science* 3234) (2004) 125-139.

[42] A. Arampatzis, M. Benkert, P. Clough, H. Joho, C. Jones, I. Reinbacher, M. Sanderson, A. Wolff, S. Vaid and M. van Kreveld, Web-based delineation of imprecise regions. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[43] R. Larson and P. Frontiera, Ranking and representation for geographic information retrieval. In: *Workshop on Geographic Information Retrieval, SIGIR 2004* (Sheffield University, 2004).

[44] D.M. Mountain and J.F. Raper, Positioning techniques for location-based services: characteristics and limitations of proposed solutions, *Aslib Proceedings* 53(10) (2001) 404-412.

[45] J. Jose and S. Downes, Evaluation of Mobile Information Retrieval Strategies. In: *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries* (2005).

[46] WebPark, *Geographically relevant information for mobile users in protected areas* (2003). Available at www.webparkservices.info (accessed 31 January 2003).

[47] K. Krug, W. Abderhalden and R. Haller, *Analysis and definition of user needs report* (WebPark, Zurich, 2002).

[48] A. Spink, B.J. Jansen, D. Wolfram and T. Saracevic, From E-Sex to E-Commerce: Web Search Changes, *Computer* 35(3) (2002) 107-109.

[49] Yell Group, *The UK's local search engine* (2005). Available at http://www.yell.com (accessed 27 April 2005).

[50] Vodafone, *Vodafone live: Hide and Seek* (2005). Available at http://www.vodafone.co.uk (accessed 28 Feb 2005).

[51] D.M. Mountain, A. Goncalves and A. Rodrigues, Geographic knowledge discovery and spatial agent interfaces for location-based services. In: *GIS Research UK* (City University, London, City University, London, 2003).

[52] D.M. Mountain, *Exploring mobile trajectories: An investigation of individual spatial behaviour and geographic filters for information retrieval* (City University, London, in press).

[53] C. Nothegger, S. Winter and M. Raubal, Selection of Salient Features for Route Directions, *Spatial Cognition and Computation* 4(2) (2004) 113-136.

[54] K. Krug and W. Abderhalden, *Report on testing and validation* (WebPark, Zurich, 2004).

[55] G.M. Giaglis, Towards a Classification Framework for Mobile Location Services. In: B. E. Mennecke and T. J. Strader (eds), *Mobile Commerce: Technology, Theory and Applications,* (Idea Group Publishing, 2003).

[56] upmystreet.com, *upmystreet.com* (2005). Available at http://www.upmystreet.com (accessed 29 Jul 2005).

[57] Google, *Google local* (2005). Available at http://local.google.co.uk/ (accessed 28 Apr 2005).

[58] Yell Group, *Yellow Pages* (2005). Available at (accessed 29 Jul 2005).

[59] J. Raper, D.W. Rhind and J. Shepherd (eds), Postcodes: The New Geography (Longman, 1992).

[60] Ordnance Survey, *1:10,000 GB gazetteer* (Ordnance Survey, Southampton, 2005).

[61] P. Fisher and J. Wood, What is a Mountain? or the Englishman who Went up a Boolean Geographical Concept but Realised it was Fuzzy, *Geography* 83(360) (1998) 247-256.

[62] P. Fisher, J. Wood and T. Cheng, Where is Helvellyn? Fuzziness of multi-scale landscape morphometry, *Transactions- Institute of British Geographers* 29(1) (2004) 106-128.

[63] T. Berners-Lee and M. Fischetti (eds), Weaving the Web: The original design and ultimate destiny of the World Wide Web (Harper, San Francisco, 1999).

[64] 3, *Introducing 3* (2005). Available at http://www.three.co.uk/aboutus/newkind.omp (accessed 28 Jul 2005).

[65] O2, *About O2* (2005). Available at http://www.o2.co.uk (accessed 28 Jul 2005).

[66] Orange, *About Orange* (2005). Available at http://www.orange.co.uk/about/ (accessed 28 Jul 2005).

[67] T-Mobile, *About T-Mobile* (2005). Available at http://www.t-mobile.co.uk (accessed 28 Jul 2005).

[68] Vodafone, *Vodafone: The company* (2005). Available at http://www.vodafone.co.uk (accessed 28 Jul 2005).

[69] VisionGain, *3G Forecasts 2003-2008* (VisionGain, 2003).

[70] Berg Insight AB, *The Structure Of The European LBS Market 2005* (Berg Insight AB, 2005).