Geo-Enabled, Mobile Services—A Tale of Routes, Detours, and Dead Ends

Christian S. Jensen

Department of Computer Science, Aalborg University, Denmark
`csj@cs.aau.dk`

Abstract. We are witnessing the emergence of a global infrastructure that enables the widespread deployment of geo-enabled, mobile services in practice. At the same time, the research community has also paid increasing attention to data management aspects of mobile services. This paper offers me an opportunity to characterize this research area and to describe some of their challenges and pitfalls, and it affords me an opportunity to look back and reflect upon some of the general challenges we face as researchers. I hope that my views and experiences as expressed in this paper may enable others to challenge their own views about this exciting research area and about how to best carry out their research in their own unique contexts.

1 Introduction

Driven in large part by rapid and sustained advances in key computing and communication hardware technologies, an infrastructure is emerging that contains vast quantities of interconnected computing and sensory devices.

Notably, we are witnessing continued improvements in the capabilities of consumer electronics such as mobile phones, personal digital assistants, personal computers, cameras, mp3 players, watches, navigation systems, and driver assistance systems. The performance and performance/price ratios associated with key technologies utilized by such systems and devices continue to increase quite rapidly.

Geo-positioning is also becoming increasingly available. For example, network assisted GPS promises to eliminate the excessive power consumption of GPS receivers, thus rendering GPS practical for outdoor, battery powered devices. The first satellite of the Galileo positioning system has already been launched, and Galileo is expected to be operational in 2010 [20]. Galileo will offer better positioning than does GPS with respect to several aspects, including the accuracy, penetration, and time to fix [1]. For example, the best-case accuracy (without the use of ground stations) of Galileo is 45 cm as opposed to 2 m for GPS. Next generation GPS will also offer better positioning, and Galileo and GPS are expected to be interoperable.

Further, the trend is towards the ability of consumer electronics devices to communicate with one another and their becoming Internet-worked.

This emerging infrastructure has the potential for enabling entirely new, geo-enabled applications and services that were either not relevant or of little use in fixed desktop computing settings.

M.L. Lee, K.L. Tan, and V. Wuwongse (Eds.): DASFAA 2006, LNCS 3882, pp. 6–19, 2006.

The range of possible applications and services is virtually limitless. For example, it includes traffic and transportation related services such as “fleet” management, including emergency vehicle dispatching and hazardous cargo and traffic offender tracking; road-pricing where payment is dependent on where, when, and how much a vehicle drives; other “metered” services, such as insurance and parking. It includes services that warn drivers about accidents, slow-moving vehicles, and icy and slippery road conditions on the road ahead. It also includes a wider range of safety-related services, such as services that track senile senior citizens, tourists traveling in potentially dangerous environments, and prisoners serving time at home. Next, it includes the oft-mentioned point-of-interest services that identify gas stations, restaurants, hospitals, etc. Finally, it includes the emerging and challenging area of games and “-tainment” (edu-, info-, enter-) services. One theme is to move games from taking part in a virtual world behind a small computer screen to taking part in reality. Virtual objects, e.g., treasures (or caches, cf. geocaching [6]), monsters, and bullets, are given geographical coordinates along with real, physical objects. This arrangement then enables games that aim to find treasures, catch or escape monsters, and hit with (virtual!) bullets.

Adopting a data centric view, I believe that by capturing pertinent aspects of reality in digital form—in semantically rich and appropriately organized structures, and with powerful update and retrieval techniques available—an ideal foundation for delivering a wide range of mobile services is obtained.

Members of the database research community are increasingly engaging in research in this exciting area, for good reasons. Geo-enabled, mobile services have great potential for being applied throughout society. Data management is a central element of such services. Further, this area offers ample new challenges to data management.

The remainder of this paper consists of four sections. In the next section, I discuss several general issues that relate to conducting use-inspired research and that reach beyond this research area. Section 3 elaborates on the data centric view of geo-enabled, mobile services espoused above. Then, Sect. 4 presents selected challenges and pitfalls specific to research within geo-enabled, mobile services. Finally, Sect. 5 summarizes the paper and points to further readings.

2 Aspects of Conducting Use-Inspired Research

This section considers first the positioning of research activities according to the degrees to which they are use inspired. Then, the positioning of research activities with respect to when their results can be expected to find practical application is discussed. Finally, Sect. 2.3 covers possible sources of inspiration for research ideas.

2.1 Solutions to Real Problems and Fundamental Insights

In his book “Pasteur’s Quadrant,” Donald E. Stokes [18] discusses the traditional dichotomy between basic research and applied research. He argues for a new, two-dimensional taxonomy. One dimension distinguishes between research that is use inspired and research that is not. The other distinguishes between research that yields (or aims for) fundamental understanding and research that does not. Stokes names the other two interesting quadrants after Bohr (fundamental insight, not use inspired) and Edison (no fundamental insight, use inspired). He gracefully leaves the last quadrant unnamed.

Our different research activities may be categorized according to Stokes' concepts. In particular, we may position our activities with respect to how use inspired they are. At the one extreme, we find application development. Here, we may well have a requirements specification that details what it takes to meet users' needs. At the other extreme, we typically find mathematical works that simply aim to solve open problems, e.g., to establish new complexity bounds for theoretical problems.

Much research in the database area—certainly in relation to the specific area considered in this paper—belongs in-between these two extremes. On the one hand, we do not base our work on articulated requirements specifications from real applications. On the other hand, we do tend to state practical concerns as motivation for our research.

Both extremes have merit, as do positions in-between these two. However, there are also dangers associated with in-between positions. In particular, when aiming for in-between positions, we run the risk of neither meeting any real needs nor solving any fundamental problems, thus ending up in the unnamed quadrant. Some years ago, I discussed this issue with a senior researcher. He told me that when he reviews papers, he is happy with a paper if it is able to simply point to one real application where its contribution is useful. At first, I thought that these were low stakes. I have since realized that this is not the case. It can actually be quite hard to identify such an application. A single paper often represents only one step towards a contribution that may have practical applications.

If we simply list, as an afterthought, specific applications in the motivational part of the introduction to a paper when the research is being written up, the results are often not convincing, and we run the risk of fooling ourselves. I believe that some of our research activities may benefit from us spending more time thinking about their positioning with respect to (specific or classes of) applications.

2.2 Timing

In many of our research activities, we aim for results that may eventually find application in practice. For this research, it seems to make sense for us to consider early on *when* we expect the results to be applied and then to formulate expectations to the state of reality as of that future time. The point is that the research results should apply to that reality.

It is of course not possible to accurately predict the state of reality, or even the aspects of reality that may be most relevant for our research, in, say, five or ten years from now. However, I still advocate that we spend a bit of time in formulating some expectations. The alternative would be to work totally in the dark.

One starting point is to extrapolate technology trends. Moore's Law effectively states that processor speeds double every 2 years (the numbers of a transistors on a chip doubles every 2 years) [8]. This self-fulfilling prophecy was put forward by Intel co-founder Gordon Moore in the mid 1960s, and it has roughly held true for four decades. Similar statements, with shorter doubling times, may be made for disk capacities and computer network bandwidths.

Let us consider what I term the Bicycle Analogy: *How fast will a bicyclist be able to go if Moore's Law applied?* Making the reasonable assumptions that the bicyclist is able to travel at 30 km/h originally and that the speed doubles every 2 years for 40 years, the current speed is 31,457,280 km/h!

This analogy illustrates several points: First, it illustrates my view that quantitative advances in hardware technologies are very important drivers of the research in software technologies. They have a profound impact on the software research agenda. (Qualitative advances, e.g., in the form of new types of sensors, are also important.) Second, sustained, exponential improvements such as these are dramatic and difficult to imagine. Indeed, they are counter-intuitive—we are simply not used to such rates of improvements in our daily lives. So, even if we have heard about exponential improvements, they are not natural to us.

I still remember how thrilled I was when, in 1991, I was able to get an external disk for my Mac. This disk was heavier than my current laptop, it sounded like a jet plane when it was turned on, and its capacity was less than 20 MB! At that time, if someone had told me about today's disks, or that I could get, e.g., a Secure Digital memory card that weighs 2 grams, uses very little power, and has a capacity of 2 GB, I am not sure that I would have believed it or acted wholeheartedly upon it in my research.

Third, the humans are the constant parts of the equation. While we perhaps think we are getting a lot smarter as the years go by, the improvements are negligible in comparison to the technology improvements.

Jim Gray's DBLP listing [5] has a significant concentration of papers that concern technology trends and is a good starting point for continued studies.

However, just because something is technically possible, this does not mean that that something is being deployed in practice. If we simply adopt a very technological focus, we may end up with overly optimistic predictions. Many technological possibilities do not materialize in practice, or do so only much later than possible. For example, third-generation mobile telephony has been technically possible for quite a few years, but is only now being deployed in many parts of the world.

People and enterprises are often conservative. The availability of existing infrastructures, or legacy systems, that to a large extent are capable of meeting needs block the deployment of new technologies. For new technologies to actually be deployed, a plausible business case must exist.

Incidentally, one big difference between academia and business is the importance of timing. In academia, it is probably not a disaster if a particular research result finds application only a few years later than expected. Rather, it is likely to be considered a success that the result found application at all! In business, where the potential financial rewards are higher, timing is of the essence. Once an enterprise has invested in new technology, that investment needs to generate revenue, so that salaries, etc. can be paid.

2.3 Inspiration

One important aspect of doing well in research is to work on great research ideas. There are many approaches to seeking inspiration that may lead to great ideas. Here, I discuss three.

A traditional way of getting ideas is to study the body of related work within one's area. In any event, it is important to be familiar with past, related work. One reason is that research results have little value if they are not new. Another is that it is important to build on past results where possible. By reading the literature, it is possible to get great ideas. However, this approach may also have a tendency to foster research activities and results that are closely related to existing activities and results. What has already been done had a tendency to define the "universe," and it is difficult to step well outside this universe.

For example, the research literature contains several dozen proposals for temporal data models and query languages. It seems that as everybody in the temporal database community read the many interesting papers about data models and query languages, they had to also study these subjects and also had to propose their own. As more and more proposals accumulated, it became increasingly harder to invent new and interesting proposals. And at some point, it probably became more constructive to study other problems.

The second approach to seeking inspiration is to interact with entrepreneurs. While researchers are constantly on the lookout for interesting problems to work on and thus thrive on problems, entrepreneurs thrive on solutions—they have plenty of problems.

During my interactions with entrepreneurs—e.g., via participation in advisory boards or boards of directors for technology start-ups, via industrial collaborations in research projects, and via participation in industry associations—many of my "great" ideas have been shot down as either not a "real" problem (because there is an easy 80% solution that does the job) or not an "important" problem (because there are many, more pressing problems).

While entrepreneurs have their own agendas and can be quite strong minded, they offer a different perspective. They serve as a filter that helps eliminate bad ideas and prioritize the remaining ones. In my experience, this has a positive effect on relevance and impact.

The third approach is to obtain inspiration from domain experts. One of the problems with relying solely on the first approach is that the research literature is generally quite abstract when it comes to the requirements of real problems. In contrast, domain experts have much richer views of problems and requirements. By interacting with domain experts, it becomes possible to "see" problems that would otherwise have remained invisible. For example, I have benefited from interacting with traffic researchers. This way, I have learned about problems that I would otherwise not have imagined.

3 Geo-Enabled Mobile Services

This section first elaborates on the data centric view of mobile services as formulated in the introduction. It then discusses the various types of content of relevance for mobile services, including business content (e.g., point-of-interest data), generic geo-content, also termed infrastructure, and user-specific geo-context.

3.1 Overview

The introduction states that "by capturing pertinent aspects of reality in digital form—in semantically rich and appropriately organized structures, and with powerful update and retrieval techniques available—an ideal foundation for delivering a wide range of mobile services is obtained." This section elaborates some on this statement.

First, this statement represents a data centric view of mobile services. The idea is that a service request by a mobile user translates into queries against the database envisioned in the statement. A key challenge in the delivery of mobile services then becomes a data processing problem.

The phrase "digital mirror of reality" has been used for describing the envisioned database. While this concise phrase is certainly to the point, it only partially reflects the desirable capabilities of this database, which go well beyond simply being a mirror. In particular, the database may capture past states of reality and one or more perceived future states, in addition to the current state. In more technical terms, the database supports the valid-time aspect of data. Further, if accountability is a concern, the database may include an incorruptible record of its past states. In technical terms, this is called transaction-time support.

Next, the database and the database management system used may not be a single relational or object-relational database stored in a centralized system. Rather, the database and system may well be distributed and heterogeneous in a number of respects. For example, the data may be physically distributed and may not adhere to the same common schema or data model. The control and data processing may also be distributed.

3.2 Infrastructure and Business Content

The delivery of geo-enabled mobile services in practice is dependent on relevant content being available. Examples of content include weather data; traffic condition data, including information about accidents and congestion; information about sights and attractions, e.g., for tourists; information about hotel rooms, etc. available for booking; and information about the current locations of populations of service users.

The management of such content includes several aspects. An information technology infrastructure, as discussed briefly in the previous section, must exist that is capable of capturing the content and capable of absorbing the content as it is made available, while being able simultaneously to make the content available to services.

We may distinguish between two types of content: the geographical infrastructure itself and all the other, "real" content that may be given geographical references and that must reference the infrastructure. So-called points of interest exemplify real content.

The *geographical infrastructure*, or *geo-content*, concerns the geographical space "itself," with hills, lakes, rivers, fjords, etc. It also concerns the road networks for use by vehicles and the transportation infrastructures for, e.g., pedestrians, trains, aircraft, and ships. The infrastructure for vehicles is of high interest because users may frequently be either constrained to, or at least using, this infrastructure.

Geo-content is essential. Users think of the real content as being located in a transportation infrastructure, and they access the content via the infrastructure. For example, the location of a point of interest is typically given in terms of the road along which it is located, and directions for how to reach the location are given in terms of the transportation infrastructure.

For the delivery of a range of geo-enabled mobile services, it is particularly important that a representation of the road infrastructure is available that supports multiple functions, including content capture; content update and querying, including route planning and way finding; and user display. This representation may be composed of several constituent representations [11].

It is common practice to specify the location of some content relative to the nearest kilometer post along a specific road. For example, the entry to a new parking area may be indicated by a road, a kilometer post on that road, and an offset. One representation, using linear referencing, is used in connection with the capture of such content.

A weighted, directed-graph representation may also be used that represents a quite abstract view of an infrastructure. This representation ignores geographical detail, but preserves the topology, and it may be used for connectivity-type queries, such as route guidance and way finding.

Next, a geo-representation is also needed that captures the geographical coordinates of the road infrastructure. With this representation, it is possible to map a location given in terms of geographical coordinates, e.g., from GPS receivers or point-on-a-map-and-click interfaces, to a location in the infrastructure. Finally, these three representations must be integrated.

All the *real content* encompasses any content that may reference, directly or indirectly, the geographical infrastructure. A museum, a store, or a movie theatre may have both a set of coordinates and a location in the road network. This type of content is open-ended and extremely voluminous. For example, it may include listings of movies currently running in the movie theatre, it may include seat availability information for the different shows, and it may include reviews of the movies. Often, the real content is the primary interest of the users.

Content is generally dynamic. This applies to road networks, where road construction and accidents change the characteristics of the networks with varying degrees of permanence. Other content is also dynamic. Examples abound. New stores open and existing stores relocate or close. The opening hours of a facility may change. The program of a movie theater changes. The sales available in a store change. This dynamicity of content implies that a representation of content must be designed to accommodate updates.

Content is more or less dynamic. The content that derives from the sampling of the positions of moving objects belongs at the highly dynamic end of the spectrum. Capturing the present positions, and possibly the past as well as anticipated future positions, of a large population of mobile users requires special techniques, as discussed next.

3.3 User-Specific Geo-Context—Locations, Destinations, Routes, and Trajectories

User-specific geo-context is another kind of content. Among such content, the *current position* of a service user is the traditional geo-context used in location-based services.

To maintain an up-to-date record of the current position of a service user, we may envision a scenario where a central server maintains a representation of the user's movement and where the local client, e.g., a mobile phone, is aware of the server-side representation. The client frequently compares its GPS position to the server-side position, and when the two differ by a threshold slightly smaller than the accuracy required, an update is issued to the server, which then revises its representation of the client's movement and sends this new representation to the client [3, 4]. This arrangement, termed shared-prediction-based tracking, aims to reduce the number of updates needed in order to main a current position at a given accuracy.

Different representations of a user's movement result in different rates of update. We consider several possible representations in turn. First, we may represent the movement of a user as a constant function, i.e., as a point. With this representation, an update is needed every time the user has moved a (Euclidean) distance equal to the threshold away from the previous position. This is a simple representation, and it may be useful when the user is barely moving or is moving erratically within an area that is small in comparison to the area given by the threshold used.

Second, we may represent the movement by a linear function, i.e., by a vector. When the user exceeds the threshold, the user sends the current GPS location and the current speed and direction (which GPS receivers also provide) to the server. The server then uses this information to predict the user's to-be-current positions.

Third, we may utilize the infrastructure in the representation of a user's movement. This requires that we are able to locate the user with respect to the infrastructure. One possibility is to assume that the user is moving at constant speed along the road on which the user is currently located. We may use the GPS speed as the constant speed, and we may assume that the user stops when reaching the end of the current road segment. Depending on the lengths of the segments, this representation can be expected to be better or worse than the vector representation. However, for realistic segments, this representation has the potential for outperforming the vector representation.

Next, we may use the route of the user in place of the segments. Folklore has it that most humans who travel do so towards a known destination. Most often, we do not move around aimlessly. Further, being creatures of habit, and perhaps for maximum efficiency, we tend to follow routes we have previously followed. Therefore, it is a good assumption that we are frequently able to predict correctly the route on which a service user travels. Using the correct route in place of a road segment means that the number of updates needed to maintain a user's position with the desired accuracy decreases further. Indeed, updates occur only because of incorrectly predicted speeds—no updates are caused by incorrectly predicted "locations." It should also be observed that if a route is predicted incorrectly, e.g., because the user makes a turn, this does not lead to a breakdown. Rather, this simply forces an update and a new prediction.

The infrastructures currently available for mobile services support the accumulation of GPS data from vehicles. Based on this data, it is possible to gradually create usage patterns for vehicles that consist of the routes traveled by the vehicles along with usage meta-data, which are temporal patterns that describe for each vehicle and route when the route is being used by that vehicle [2]. For example, a pattern may specify that a route is being used in the morning on weekdays. The resulting route and destination data may subsequently be used in services. By also attaching travel speeds to routes, we obtain trajectories, which are routes "lifted" into the time dimension [7].

4 Pitfalls and Specific Recommendations

This section presents six recommendations for conducting research. These are intended to apply to the area of geo-enabled, mobile services, but are to varying degrees applicable also to other areas of research. These are all recommendations that I am trying to follow myself, with the hope that my research is going to benefit.

4.1 Perceived Reality

> *For application-oriented research, estimate the time of application and formulate expectations to the reality as of that time; then design for that reality.*

This recommendation was discussed in an abstract setting in Sect. 2.2. My different research activities[1] have quite different use horizons. For example, my research on tracking (e.g., [3, 4]) is applicable here and now, and is expect to remain applicable for the foreseeable future. For this research activity, we take care to only make assumptions that are met by current infrastructures. Assumptions concern the available computing and storage capabilities of mobile terminals, the available communication technologies, the available positioning technologies, the available digital road networks, and existing legislation.

Towards the other extreme, I expect that much of my research on the indexing of the positions of moving objects (e.g., [12, 14, 15, 16, 17]) is only applicable in the longer term. For disk-based indexing to be of interest, the sets of data items to be indexed must be much larger than the data sets seen today. For indexes that consider the current states of objects, the data sets should probably contain positions of hundreds of thousands of objects, while for indexes that consider the entire lifetimes of objects, data sets that concern thousands or tens of thousands of objects suffice to render disk-based indexing relevant.

This line of research is more speculative than the research on tracking, and it is also somewhat more removed from specific applications. Some of the results may not offer the final answers, but may serve as inspiration for further work. Also, although this research is generally cast in the setting of indexing of moving object, it might be that the results will be applied in other settings, e.g., settings with low-dimensional, continuous variables.

4.2 Architectural Setting

> *Ensure that at least one appropriate architectural setting exists or may be envisioned for the research contribution.*

For some research, it is important to be specific about the architectural assumptions underlying the research. For other research, it may be sufficient to ensure that an appropriate architectural setting exists or can be envisioned. And for yet other research, architectural settings may not be an important concern.

In particular, for research that is expected to have practical application in the short term, the architectural setting is likely to be a concern. In keeping with this, the research I have conducted with my colleagues on tracking and also route acquisitioning and provisioning [2] is fairly explicit about architecture, and attempts have been made to ensure

---

[1] I will generally concern myself with my own research, to avoid making bold statements about the research of others.

that the assumptions about the possible application contexts are reasonable. In particular, we believe that the contributions are applicable in current application contexts, e.g., existing server-side systems, existing mobile terminals that use GPS for positioning and for GPRS data communication, and existing digital road networks. To justify these claims, we have built and demonstrated proof-of-concept prototype implementations.

Considering next the research on the indexing of moving objects, which is less directly applicable in practice and only in the longer term, the issue seems to be to ensure that architectural settings will indeed exist at the time when the various proposed indexes become widely applicable.

One point here is that it appears unrealistic to assume that the many indexes for moving objects will find their way into conventional object-relational database management systems. Other areas of data management and computer science research are also quite prolific when it comes to the invention of new indexes, so these observations apply also to those other areas. We may instead assume that the indexes may be applied in more componentized and open data management architectures.

4.3 Composability

Invent solutions for composable functionality.

When research on query processing in relation to moving objects first took off, the efficient processing of many basic types of queries had yet to be explored in the new moving-object settings. Examples included one-time and continuous range queries, nearest-neighbor queries, and reverse nearest-neighbor queries, to name but a few.

As techniques for the processing of these basic types of queries accumulate, it is natural that attention shifts to as yet unexplored or lightly explored types of queries. A potential pitfall is that we start producing highly optimized solutions to very specialized types of queries. This path is not advisable, as the prospects of these solutions finding practical applications are likely to decrease with the degree of specialization of the functionality.

To appreciate the point, consider SQL and the relational algebra as examples: We should avoid following the path where we invent highly efficient algorithms for increasingly complex SQL queries. Rather, we should focus on developing efficient algorithms for the relational algebra operators in terms of which the SQL queries may be expressed. At some point, query optimization should take over from efficient, standalone algorithms.

4.4 Versatility and Robustness

Prioritize versatile and robust solutions over specialized and brittle, although possibly highly performant, solutions.

One lesson to be learned from current, commercial data management technology and existing applications is that versatile and robust solutions have better chances of finding practical use than do very specialized ones, even if these exhibit very high performance in some cases. The objective of a query optimizer is quite modest: it should avoid the clearly inefficient ways of computing a query and identify one good way of executing the query. (Even meeting this objective can be a challenge.)

In the area of data processing for moving objects, the parameter space—the number of parameters and parameter settings that characterize a data processing workload—is very large. One consequence of this is that there is "room" for many solutions that offer superior performance for certain settings, but may be clearly inefficient for other settings.

The recommendation is that we try to aim for solutions that are versatile in terms of the functionality they offer and that are robust in terms of the settings. A solution for which there exist other solutions so that for every possible parameter setting, at least one of these other solutions has twice the performance may still be preferable if it is much more robust than its competitors.

One concern here is that is seems to be much easier to produce an experimental study that demonstrates the merits of a highly performant, but possibly brittle, solution than a study that demonstrates the merits of a robust solution with performance that is dominated by existing solutions.

4.5 Context

Design query processing techniques that exploit the entire context.

Mobile services are delivered to devices that are typically without (qwerty) keyboards and that have only small screens. Further, a service may be expected to be delivered in situations where the main focus of attention of its user is not the service, but rather that of, e.g., navigating safely in traffic. For these reasons, it is much more important than in a desktop computing situation that the user receives only the relevant information and service, with as little interaction with the system as possible. One approach to obtaining these qualities is to make the mobile services aware of the user's context, as covered in Sects. 3.2 and 3.3. Another benefit of taking the entire context into account is that better functionality can be provided.

The user's current location is one possible geo-context, and the user's destination is another. Yet another is the route that takes the user from the current location to the destination. Also, the trajectory that takes the user to the destination is a possible context.

Routes are interesting for at least two reasons. First, as also discussed in Sect. 3.3, mobile users typically travel towards destinations. A user often, or typically, follows the same route when going from one location to another. For example, a user typically travels along the same route from home to work.

Second, routes are significant as context for a range of services. For example, a service that knows the route of a user may alert the user about travel conditions, e.g., congestion and accidents, on the route ahead, while not bothering the user with conditions that do not relate to the user's route. As another example, routes may be used when users request the locations of "nearby" points of interest.

Another type of geo-context is the infrastructure, e.g., the transportation infrastructure, into which the users are embedded.

When we design query processing techniques, I recommend that we try to use as arguments all the context that we can reasonably expect to have available. So if we can assume to know the likely route of a moving vehicle, we may suggest restaurants or gas stations to the driver that are near to the expected route, rather than merely to the driver's current location, which is the best a service can do if it ignores the route. And, utilizing knowledge of the road network, we can use network distances as opposed to Euclidean distances in our calculations, and we can augment the answers with distances, detour distances, and suggested routes to the points of interest returned.

4.6 Queries and Updates

Pay attention to both query performance and update performance.

Many indexing and query processing techniques for geographical data were originally developed for largely static data. For example, R-trees do not contend well with workloads with frequent updates.

In contrast, mobile-service application scenarios exist that are characterized by frequent position updates. This puts focus on techniques that are capable of supporting workloads consisting of frequent updates as well as queries, and it puts focus on studies of the trade-off between query performance and update performance.

Updates of moving-object positions correspond to the sampling of continuous, position-valued variables. One implication of this is that our record of the position of a moving object is inaccurate. Different services may tolerate different inaccuracies. For example, a localized-weather service may tolerate a relatively high degree of inaccuracy without this affecting the functionality of the service, while a navigation service is dependent on more accurate positions.

An obvious approach to taking advantage of the different accuracy tolerances of different services is to perform updates only when needed to maintain the accuracies needed (cf. Sect. 3.3). Indexing and query processing techniques should be able to exploit this approach to updates.

By forming predictions of the future movements of the objects, the numbers of updates can be further reduced. Indexing techniques for moving objects that represent the current and near-future positions of the objects as linear functions from time to points in space predict that the objects move in linear fashion. Techniques that represent object positions as points in space predict that the objects do not move. One study of the movements of vehicles [3, 4] shows that constant prediction leads to almost three times as many updates as does linear prediction for a range of reasonable accuracies.

5 Summary

Based on my own research experience and with a focus on my research in the area of geo-enabled mobile services, this paper first presents some of my general thoughts about conducting use-inspired research. Following a data centric characterization of geo-enabled mobile services and the content of relevance to such services, the paper presents six recommendations for future work in mobile services.

Although I try to maintain a portfolio of research activities that range from ones with practical applicability in the short term to ones that are more speculative and that may only have indirect applications in the long term, the paper mainly concerns research with intended applications in the near and medium terms. Thus, those who conduct research in an abstract setting that is unrelated to perceived applications may not find the paper relevant.

It is important to realize that there is no single best approach to obtaining good research results. I hope that the thoughts presented in this paper can inspire others to possibly adjust the approaches they favor, so that they avoid detours and dead ends in their research and instead are able to identify direct routes to even better results.

For those who are interested in introductions to the general area covered in this paper, the recent books by Voisard and Schiller [19] and Güting and Schneider [7] come highly recommended. Reaching beyond data management, the first offers a broad coverage of location-based services, while the second is devoted specifically to data management for moving objects. Also, two recent special issues [9, 10] of the IEEE Data Engineering Bulletin are good starting points for those interested in doing research in indexing and query processing for moving objects.

Acknowledgments

I would like to thank my current Ph.D. students and colleagues with whom I conduct research on geo-enabled mobile services: Agnė Brilingaitė, Alinas Čivilis, Xuegang Huang, Stardas Pakalnis, Torben Bach Pedersen, Simonas Šaltenis, Albrecht Schmidt, Laurynas Speičys, Dalia Tiešytė, Kristian Torp, and Nerius Tradišauskas. Thanks also to my other recent collaborators within this paper's area: Rimantas Benetis, Michael Böhlen, Martin Breunig, Bin Cui, Anders Friis-Christensen, Ralf Hartmut Güting, Christian Hage, Zhiyong Huang, Wynne Hsu, Gytis Karčiauskas, Augustas Kligys, Jan Kolář, Harry Lahrmann, Mong Li Lee, Dan Lin, Hua Lu, Jovita Nenotaitė, Beng Chin Ooi, Dieter Pfoser, Rosanne Price, Keng Lik Teo, Igor Timko, Nectaria Tryfona, Rui Zhang, and Nora Zokaitė.

It is in large measure my interactions with these dedicated and capable individuals that have led me to formulate the views expressed in this paper (but for which I am solely responsible!).

References

1. BBC News: Q&A Europe's Galileo project. news.bbc.co.uk/1/hi/sci/tech/4555276.stm(2006)
2. Brilingaitė, A., Jensen, C.S., Zokaitė, N.: Enabling routes as context in mobile services. In: Proc. ACM GIS (2004) 127–136
3. Čivilis, A., Jensen, C.S., J. Nenortaitė, J., Pakalnis, S.: Efficient tracking of moving objects with precision guarantees. In: Proc. MobiQuitous (2004) 164–173
4. Čivilis, A., Jensen, C.S., Pakalnis, S.: Techniques for Efficient Tracking of Road-Network-Based Moving Objects. IEEE TKDE **17** (2005) 698–712
5. DBLP: Jim Gray. www.informatik.uni-trier.de/~ley/db/indices/a-tree/g/Gray:Jim.html (2006)
6. Groundspeak: Geocaching. www.geocaching.com (2006)
7. Güting, R.H., Schneider, M.: Moving Objects Databases. Morgan Kaufmann (2005)
8. Intel Corporation: Moore's Law. www.intel.com/technology/silicon/mooreslaw/ (2006)

9. Jensen, C.S. (ed): Special Issue on Indexing of Moving Objects. IEEE Data Engineering Bulletin **25(2)** (2002).
10. Jensen, C.S. (ed): Special Issue on Infrastructure for Research in Spatio-Temporal Query Processing. IEEE Data Engineering Bulletin **26(2)** (2003).
11. Jensen, C.S.: Database Aspects of Location-Based Services. In: [19] (2004) 115–147
12. Jensen, C.S., Lin, D., Ooi, B.C.: Query and update efficient B+-tree based indexing of moving objects. In: Proc. VLDB (2004) 768–779
13. Jensen, C.S., Šaltenis, S.: Towards Increasingly Update Efficient Moving-Object Indexing. In: [9] (2002) 35–40
14. Lee, M.L., Hsu, W., Jensen, C.S., Cui, B., Teo, K.L.: Supporting frequent updates in R-trees: a bottom-up approach. In: Proc. VLDB (2003) 608–619
15. Lin, D., Jensen, C.S., Ooi, B.C.: Efficient indexing of the historical, present, and future positions of moving objects. In: Proc. MDM (2005) 59–66
16. Pelanis, M., Šaltenis, S., Jensen, C.S.: Indexing the past, present and anticipated future positions of moving objects. ACM TODS **31** (2006, to appear)
17. Šaltenis, S., Jensen, C.S., Leutenegger, S.T., Lopez, M.A.: Indexing the positions of continuously moving objects. In: Proc. ACM SIGMOD (2000) 331–342
18. Stokes, D. E.: Pasteur's Quadrant. Brookings (1997)
19. Voisard, A., Schiller, J.: Location-Based Services. Morgan Kaufmann (2004)
20. Wikipedia: Galileo positioning system. en.wikipedia.org/wiki/Galileo_positioning_system (2006)