

Involving LBS users in data acquisition and update

Florian Sayda
GIS lab (AGIS) University of the Bundeswehr Munich
Werner-Heisenberg-Weg 39
85577 Neubiberg, Germany
florian.sayda@unibw-muenchen.de
Tel. ++49 89 6004 2597; Fax. ++49 89 6004 3906

SUMMARY

One of the most important components of Location Based Services (LBS) is geodata. Lots of services providing information to the LBS users are based on this kind of data. For the content and LBS providers it is very important to deliver up-to-date, accurate and reliable information. However keeping a database for LBS that satisfies the just mentioned criteria all the time is a time consuming and consequently costly matter. In this paper an approach is presented to support the content and LBS providers in this task by involving the LBS users in the data acquisition and update process for attribute information as well as spatial data. Therefore an approach with two major steps, data gathering and data evaluation, is introduced. Within the data gathering process multiple information from lots of different users is collected. Using this data the evaluation process should contribute to preventing insertion of false information into the database. To achieve this the presented approach proposes a comparison of the multiple data gathered by different user to derive correct and reliable information by retrieving and rating corresponding information. Depending on the kind of data captured by the users different problems arise. Gathering textual information results mainly in semantic problems. The capturing of spatial data and the special challenges met here are discussed with respect to the presented approach on a LBS for hikers and mountaineers in which the users should be involved in the update process of the trail network (spatial data).

KEYWORDS: LBS, data acquisition, spatial data, update, trust, reputation

INTRODUCTION

From the conceptual and technological point of view Location Based Services (LBS) are well investigated and some are already running on a commercial bases. There are initiatives defining core services and protocols within LBS, like the OpenLS specifications of the Open Geospatial Consortium (OGC 2004) or the Mobile Location Protocol (LIF 2002, OMA 2004) from Open Mobile Alliance (OMA). Although several LBS providers offer different kinds of service to potential users, the major commercial benefits have not taken place yet as predicted. Within the final report of the EC funded study "Digital Content for Global Mobile Services" (Andersen 2002), the following seven major forces necessary to stimulate the mobile market have been identified:

- Availability of the content
- Provision of relevant applications
- Awareness of users on the benefits of mobile content
- Development of a handset installed base
- Adequate revenue sharing between players
- Stimulation of investment in all enabling technologies
- Promotion of the uniqueness and inter-operability of standards

The first point mentioned in this list, the availability of the content, is with respect to LBS data. Data are one of the most important components of LBS, as all information provided to the user is derived from them. Therefore the LBS users have the following general requirements on the provided information, which are also directly applicable on the data:

- Up-to-dateness
- Completeness
- Correctness

Considering the enormous amount of data and the large areas that can be covered by LBS from the content providers point of view fulfilling these requirements is a time-consuming and costly matter. Anyway these requirements have to be met to ensure the users are satisfied with the offered services. A user receiving wrong information will not use the service in the future anymore.

In the following an approach is presented for supporting the content provider in the acquisition and update of the data by involving the users in this process. At first the approach is explained in general. Later on it is applied on a special use case in which the users are involved in the update of geodata (i.e. trail network).

GENERAL APPROACH FOR DATA ACQUISITION AND UPDATE BY INVOLVING THE USERS OF A LBS

The approach presented in this paper to involve the users of LBS in the data acquisition and update process is separated into two major steps. The two-staged approach has been chosen, as it is not reasonable to allow every user to make changes to the database as this would very quickly result in inconsistencies within the data. The major intent of step 2 (i.e. evaluation) is exactly preventing this by comparing data on the same objects gathered by different users and thereby deriving correct and reliable information to be inserted into the database by retrieving and rating corresponding information.

The first step “data gathering” is dealing with how the user can be involved in the data gathering process and is separated into the following three sub-steps:

- 1.1 Gathering of data by the users: Within this step the user has the possibility to gather information through an appropriate interface on a mobile device (client). Depending on the kind of data to be gathered different software and hardware have to be utilized. Attention should be paid to no additional hardware besides the one necessary to use the LBS is required for the gathering of the data. This allows for enabling lots of users to participate in this process, as they do not need to buy additional hardware. This is an essential requirement for the presented approach. Concluding this sub-step the gathered data are transmitted to a server.
- 1.2 Preprocessing of the data: Once the gathered data from sub-step 1.1 are available on the server a preprocessing might be necessary. Depending on the kind of data this can be plausibility or consistency checks. In case of geodata, e.g. a polyline, this can also include reduction of point density using generalization algorithms (further explanations on this topic will be given later on)
- 1.3 Storing the data on the server: Concluding step one the preprocessed data gathered by the users are stored for further usage within step 2.

Within the second step “data evaluation” all data gathered by the users are compared so as to derive relevant and correct data. Purpose of this step is to judge the reliability of the data and hence to ensure that wrong data are not entered into the database of the LBS. The basic idea is, simply speaking, that it is more likely that a dataset is correct if it is equally gathered independently from different users. Within the approach this concept is called trust in the data gathered by the users. To achieve this the second step is divided into the following three sub-steps:

- 2.1 Extraction of data on the same objects: Within this sub-step among all data gathered by the users the ones on the same object have to be determined. Depending on the kind of data different methods have to be utilized to achieve this.
- 2.2 Comparison and rating of the data on the same object: Data on the same object determined in the previous sub-step have to be checked for correspondence. In case corresponding data are found the data confirmed by different users have to be rated. The objective here is to evaluate the trust in the data as explained earlier. This can for example be achieved by considering the number of users who gathered the corresponding data on the same object. A simple approach here could be to allow only data to be entered into the database of the LBS if different users gather them identically. In recent years more sophisticated concepts for modeling trust often also called reputation within IT-systems have been introduced. One of the first can be found in (Marsh 1994). An overview of some of the algorithms applied in this context can be found in (Westermaier 2000) or (Zacharia 2000). The usage in the classic fields of application for these concepts like online-communities or virtual market places is described in (Grandison 2000).
- 2.3 Update of the data base: The data fulfilling the criteria for the trust in the data - in the simple example in the previous sub-step; this would be that the data are verified by a certain number of different users – are used for updating the data base of the LBS.

Different kinds of data

The approach presented in the previous section for data acquisition and update by involving the users of LBS is independent from the kind of data. Besides the kind of data has some impact on the procedures within several sub-steps. This includes the way the user gathers the data as well as the preprocessing of the data and the extraction of data on the same object. If the users participate in the acquisition and update process of textual information (e.g. attributes) on existing objects, in most cases, it is quite easy to derive the desired information required in sub-step 2.1 as these objects normally have a unique identifier. The comparison of different textual data on the same object (sub-step 2.2) to find out the ones with equal meaning is related to semantics and lots of research have been done and are currently undertaken.

If the users gather geometric data, e.g. new Points of Interest (POIs), finding data from different users on the same object (sub-step 2.1) is not that quite easy, as the geometric data (i.e. coordinates) captured by different users will not be exactly the same. As commonly known the comparison of geometric data can only take place using an appropriate buffer. The size of this zone strongly depends on the method, e.g. terrestrial surveying or Global Navigation Satellite System (GNSS), used to determine the coordinates. In the following the approach described earlier will be applied for trails to be captured by the users within a LBS for hikers and mountaineers. In the context of this subsection the trails represent geometric data in the shape of polylines.

PARAMOUNT- A LBS FOR HIKERS

This paragraph contains a description of a prototype for a LBS for hikers and mountaineers. It provides the frame in which the approach presented in the previous chapters is applied on the trail network for research purposes.

Service description

Within the EC funded project PARAMOUNT (Contract N° IST-2000-30158) a prototype for a Location Based Service for hikers and mountaineers has been developed and tested in areas in the Alps and Pyrenees. The project has been carried out by:

- IfEN GmbH, Poing
- AGIS UniBW Munich
- ICC, Institut Cartogràfic de Catalunya, Barcelona

- Bergwacht Bayern (Bavarian Mountain SAR)
- Österreichischer Bergrettungsdienst (Austrian Mountain SAR)

Main objectives of the mobile services are provision of information to the users and increase of their safety. The following main functionalities are available within PARAMOUNT:

- Visualization of the users position on a map
- Routing
- Guidance of the user along a route
- Provision of various kinds of information, e.g. weather or tourist information
- Emergency call and support of SAR missions

Figure 1 shows the overall PARAMOUNT system architecture with the three main service categories INFOTOUR, SAFETOUR and DATATOUR.

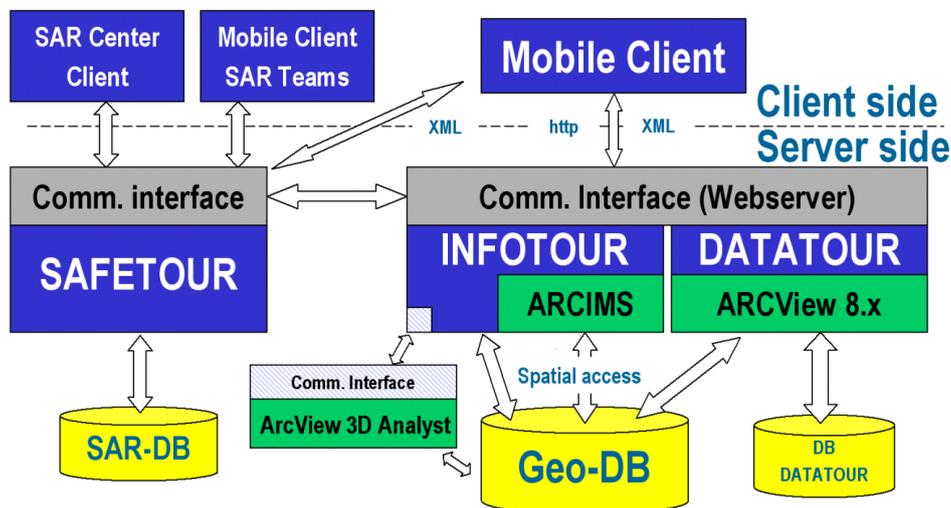


Figure 1: PARAMOUNT system architecture

As mentioned earlier content (data) is one of the most important components within a LBS. Within PARAMOUNT the following data are used:

- Raster Data (e.g. topographic maps)
- Factual information, e.g. information on POIs, weather information, ...
- Vector data
 - Points of Interest
 - Trail network

This paper deals especially with the challenges arising from applying the proposed approach on the update of the trail network by involving the users in this process, as the trail network is an essential part of the database. It cannot be easily acquired and updated. Experience shows that existing hiking maps are often old and inaccurate. Digitizing the trail network from remote sensing data is also not possible in all cases as the resolution available nowadays is too large to recognize narrow hiking trails. Considering that the costs for data acquisition have to be kept low to be able offering the services at a reasonable and affordable price, all other methods, e.g. terrestrial surveying, are too expensive. Therefore the proposed approach offers a promising solution for this problem. Further

information on the PARAMOUNT project can be found at (Reinhardt et al. 2003) or the project webpage www.paramount-tours.com.

TRAIL ACQUISITION AND UPDATE WITH LBS USER INVOLVEMENT

Within this section the presented approach is applied on involving LBS-users in the acquisition and update process of a trail network. For applying the approach the following preconditions are made:

- The geometry of the trails is represented by polylines, which is a set of points connected by straight lines. This holds true for the trails in the database that are used to test the presented approach as well as the routes captured by the users. The difference between those two is that trails in the database compose a topological network in which each trail is connecting two nodes. The routes represent the hikes of the users with no topological properties.
- The routes are captured using low-cost commercially available GPS receivers with nearly the same accuracy. The positioning accuracy of such devices is normally around below 10-20 meters, which is similar to the accuracy achieved by digitizing commercial off the shelf available hiking maps.
- The routes do not contain any major errors, like positioning errors due to lack of a good constellation of satellites. This can be achieved by applying appropriate methods during the preprocessing step (1.2) of the data.
- The sources, meaning the users, of the captured hikes are known, e.g. by a unique id.
- The users do know they should only use valid hiking trails.

In the following the main focus will be put on step 2, especially on substep 2.1 and 2.2, of the proposed approach in regard to involving the users in the update of a trail network. The input for these substeps is a database containing the hikes captured by the users as described in the general approach in step 1. This includes preprocessing in this context meaning simplification (e.g. point density reduction) and removal of large errors in the geometry of the hikes caused by GPS usage. The objective now is to derive trails, which do not run cross-country from this collection of hikes. As geometry is involved here it is most important to distinguish between the geometric accuracy of the captured hikes and the reliability of the data. In this paper these two issues are considered independent from each other. The reliability of the data should indicate if the user has really used a valid trail and has not been hiking cross-country. Obviously this can be influenced directly by the user by simply using valid hiking trails. Contrary to this the geometric accuracy of the captured hikes is considered not to be influenced directly by the users. It is mainly influenced externally by the positioning accuracy of GPS.

Within the two substeps two major research topics have been identified and will be addressed in the following:

1. As described in the general approach substep 2.1 requires, the determination of same objects gathered by different users. In the context of the trails the first question to be answered is: What is an object? To answer this the following two characteristics of the trails captured by the users have to be considered:
 - The trails are represented by a polyline (i.e. a set of points connected by a straight line)
 - Due to the positioning accuracy of the low cost GPS receivers and of the users not always hiking in each others foot steps when using the same trail the captured trails will not match exactly (see Figure 3), but only within a specific tolerance.
 - As users are allowed to submit whole tracked hikes these tracks will not match over the whole track. To users my only use partially the same trails as the might have different destinations or prefer other routes.

Therefore only parts of the trails captured by the users, called sections in the following, will represent parts of trails hiked by different users (see Figure 3). These sections represent the “same objects” as required by the general approach in substep 2.1.

The geometric algorithm for determining the same sections from a set of trails captured by the users and for calculating the resulting trail (substep 2.2) has to fulfill the following two requirements. Firstly, it has to be capable of determining the same sections of trails captured by the users within a given buffer range, e.g. 15 meters. Secondly the algorithm has to calculate new sections from the sections determined before. During this calculation the algorithm has to be able to consider different weight of the input sections.

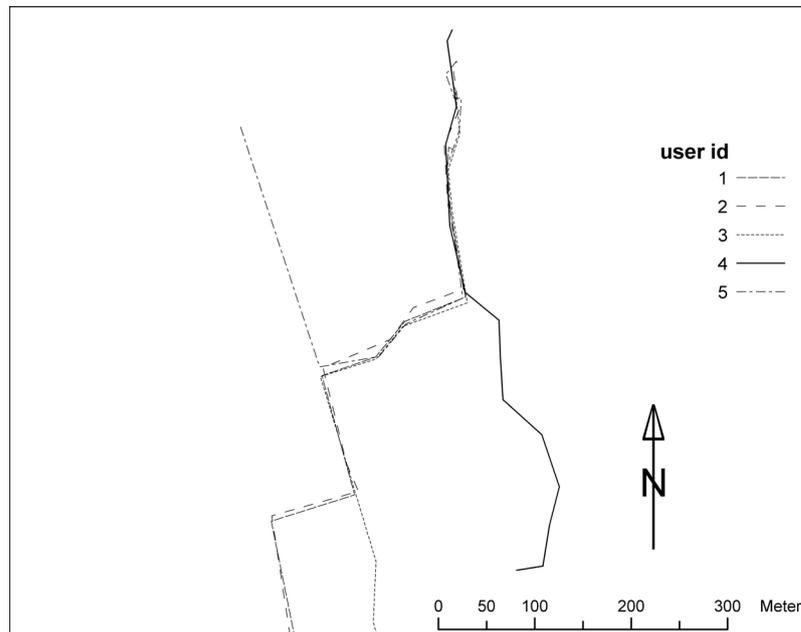


Figure 2: 5 trails from different users (1-5) with matching sections

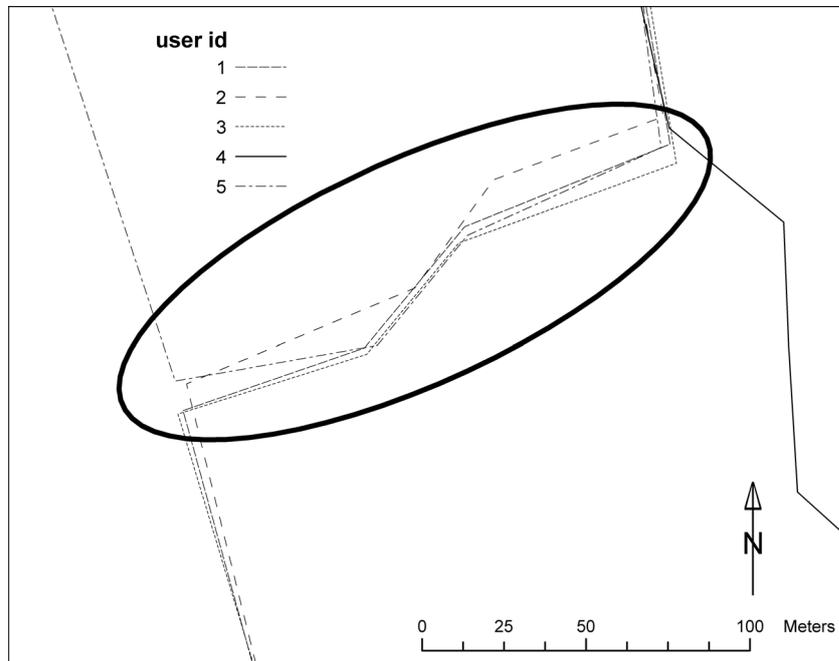


Figure 3: Equal section of different trails hiked by all 5 users

The geometric algorithm used within the test of the general approach on trails captured by users is briefly described in the following. A completed description in detail is not possible within this paper.

1. From all available trails captured by the users two are determined having likely common sections within a given buffer zone (i.e. tolerance).
2. From all the points of each trail, the shortest perpendicular to the other trail is calculated and the valid ones are determined. The perpendiculars have to fulfill several requirements to be considered valid. The most important requirements are that the length of the perpendicular must not exceed the defined tolerance and they are not allowed to intersect with any involved trail.
3. From the set of valid perpendiculars the new calculated section is determined by defining the points of the resulting trail on the perpendiculars. Depending on the weight of the input trail the points of the resulting section will have a different position on the perpendicular. In case both input trails have the same weight the resulting points will be in the middle of the perpendicular. The necessity for considering weights of the input trails will become clear if the reliability of the users is introduced in the next chapter. Figure 4 shows two examples of the results of the geometric algorithm, the input trails and the used perpendiculars.

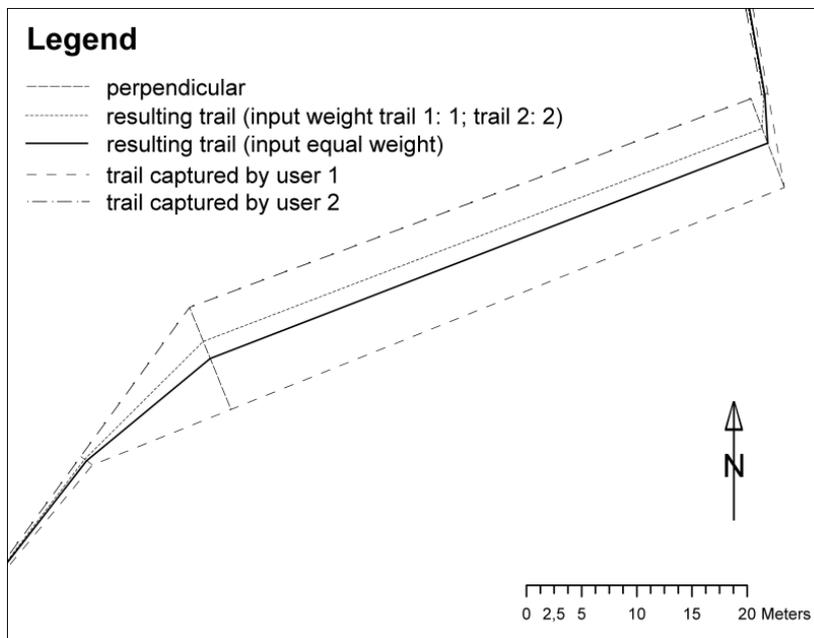


Figure 4: Two trails captured by users, resulting trail and perpendiculars

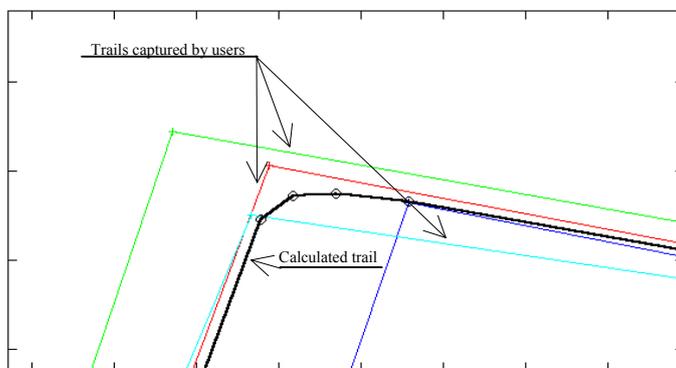


Figure 5: Example for merged trails (black) from routes delivered by users

2. Furthermore, a method has to be found to solve the question like how to rate the new calculated sections from the hikes delivered by the users (substep 2.2). The crucial question in this context, which should be answered by introducing the concept of trust in the data of the users, is “Did the user capture a route using hiking trails?”. It is not desired to insert cross-country running trails as hiking trails in the database, as it is not intended to send other users along such cross-country trails also.

The main objective here is to get a rating for the sections from the previous step, based on comparison to a threshold, indicating if the new calculated section can be inserted into the database. Simply speaking the chosen approach can be summarized like this: “If enough

users say this section is a valid trail, it is likely that it really is a valid one". One approach to handle this could be to count the number of users having captured the same section and inserting the section into the database of the LBS, if a predefined threshold is reached. The concept of trust and reputation found in IT-systems as described earlier offer a more sophisticated approach. Compared to the simple "count approach" they are trying to give a numerical value for the current reputation of an agent based on e.g. the current reputation of the rating user and the current reputation of the rated user. The agent in this context does not necessarily have to be a person. These concepts are also used for example for nodes of networks or software modules.

Within the proposed approach first tests with a method based on the Sporas algorithm (Zacharia 2000) have been conducted to get an indicator for the reputation of the users with respect to the delivery of geodata – here trails. The indicator is used within the algorithm to rate the equal trail sections derived in the previous step as well as rating the user. As the approach is an iterating process meaning the new calculated section are inserted in the database as hikes again, like the hikes captured by the users, they accumulate a reputation until the new calculated section reaches a reputation threshold indicating that they can be considered as trusted, meaning inserted into the database. In other words, a trail is considered to be trusted if it has been hiked and captured by different users having a good reputation and hence this trail can be inserted into the database. The other way round the reputation of users contributing to trusted trails will be increased at each contribution.

This iteration process is the reason why the geometric algorithm has to be able to consider weights of the input trails. For example this allows for assigning a section calculated already from two hikes more weight (2) on the position of the new calculated section if it is merged with a section (weight 1) from a new hike (see Figure 4).

First test with this approach show that the advantages offered by the reputation concepts as used in IT-system can be applied to the presented approach. Though still research is necessary here. One topic in the future will be how the rating, in the case of Sporas a numerical value from 0.1(bad) to 1 (excellent) for a specific action necessary can be calculated automatically. In some fields of application of those algorithms the human beings participating in the interaction do this. In the presented approach research is still made on how this rating can be automatically calculated by the algorithm deriving the corresponding sections from the hikes captured by the users.

CONCLUSION

Intensive research has been done on the proposed approach and has been tested in first prototypes. The first results of this tests show that it is possible to derive equal trails from the routes delivered by multiple users using the proposed geometric procedure. Also the concept of trust in data and reputation of the users in the context of geodata has been tested with a simple method and proved quite feasible. Further research has to be made on some refinements of the geometric process and especially on the modeling and implementation of the trust in the data and reputation of the users in the context of geodata concept.

BIBLIOGRAPHY

- Andersen: Digital Content for Global Mobile Services. Final Report. European Commission Directorate-General Information Society, February 2002
- Grandison, Tyrone; Sloman, Morris: *A Survey of Trust in Internet Applications*. IEEE Communications Surveys & Tutorials, 4th Quarter, 2000
- Location Inter-operability Form (LIF): *Mobile Location Protocol Specification*. 6 June 2002, <http://www.openmobilealliance.org/tech/affiliates/lif/lifindex.html> (last visited 02.03.2005)
- Marsh, S. P.: Formalising trust as a computational concept. Ph.D. Thesis, University of Stirling, Stirling, Scotland, U.K., 1994

- Open Mobile Alliance (OMA): *Mobile Location Protocol (MLP)*. Candidate Version 3.1- 16 März 2004, http://www.openmobilealliance.org/release_program/docs/MLP/OMA-LIF-MLP-V3_1-20040316-C.pdf (zuletzt besucht 16.03.2005)
- Open GIS Consortium Inc.: *OpenGIS[®] Location Services (OpenLS): Core Services*, Copyright © 16 Januar 2004 Open GIS Consortium, Inc. <http://www.opengis.org/>. All Rights Reserved. <http://www.opengis.org/legal/>. 155 pages, January 2004
- Reinhardt, W. ; Sayda, F. ; Kandawasvika A. ; Wang, F. ; Wittmann, E.: Geoinformation and Location Based Services - Principles and some Experiences with Applications for Mountaineers. In *GeoBIT/GIS Magazine*", 6/2003, Herbert Wichmann Verlag, Heidelberg; June 2003
- Westermaier, T.: *Trust in Bots: Zur Rolle von Vertrauen und Reputation bei Multiagenten-Systemen*. Studienarbeit, Albert-Ludwigs-Universität Freiburg im Breisgau, Institut für Informatik und Gesellschaft, Abteilung Telematik, Lehrstuhl Prof. Dr. Günter Müller, August 2000
- Zacharia, G.; Maes, P.: Trust Management Through Reputation Mechanisms. *Applied Artificial Intelligence*, 14, Page 881-907, Taylor & Francis, 2000.