Integration of Smart Components for Building Your Personal Mobile Guide

Rainer Malaka, Robert Porzel, Alexander Zipf and Vasu Chandrasekhara
European Media Laboratory
Schloß-Wolfsbrunnenweg 33
D-69118 Heidelberg

Abstract. Intelligent mobile systems such as Deep Map, a mobile tourist guide, require a range of intelligent components in order to actually go beyond classical systems. These components cover smart user interfaces, designed for mobile use and incorporating the use of natural language. They also include services that are location dependent, make use of context information and that allow access to a variety of different sorts of information. In our framework research from the fields of geographic information systems, natural language processing, information integration and visualization is combined in a user-adaptive mobile and web-based prototype. The central aim of the entire undertaking is to create systems that offer real users a broad range of non-trivial services in an intuitively usable way.

1 Introduction

Information technology is rapidly moving small computerized consumer devices and hi-tech personal appliances from the desks of research labs into the sales shelves and our daily life. This technological trend from main frame computers via personal computers to portable and wearable computers will lead ultimately to the disappearing computer where people will be surrounded by intelligent things that are mutually connected by wireless communication networks and provide an information infrastructure that can be used everywhere. Such a vision of a the future information infrastructure could be compared with the availability of electric light today. We are all used to the presence of electric light in our daily life such that we do not even have to think where its power is coming from and how it actually works. Whenever we enter a room at night, we can turn on the light. In some years, we might have a similar seamless information access in our daily life. We can enter any room or car and just ask questions like ”how do I get to the airport” or ”what is the German word for *negotiation*” and our information infrastructure, e.g., our wrist computer, car computer, personal network, mobile phone etc., will find an answer and a way to present the result to us.

Today, however, we are still far from this scenario. Even though hardware offers more computational power in smaller devices and networking technology becomes available for all sorts of devices, there is still a lack in intelligent software that drives these new gadgets. Various platforms from low performance PDAs, embedded computers in cameras, cars, or mobile phones, up to high performance wearable computers require new interaction metaphors and methods of control. Resources such as power or networking bandwidth may be limited depending on time and location. Moreover, the physical environment and context can change rapidly and must be taken into account appropriately.

The main challenge for the success of mobile systems is the design of smart user interfaces and software that allows ubiquitous and easy access to personal information and that is flexible enough to handle changes in the user's context and the availability of resources. Artificial intelligence has investigated the problems of making user interfaces smart and cooperative for many years and is attacking the challenges of explicitly dealing with limited resources lately. AI methods provide a range of solutions for those problems and currently seem to be the most promising tools for building location and situation aware mobile systems that support users at their best and behave cooperatively in unobtrusive ways.

In Deep Map, a mobile tourist guide, we address some of these questions. We developed a prototypical architecture and implementation that involves

- a wearable computer as a mobile platform,
- a network of databases managing geographical and touristic information,
- the employment of multi-modal user interfaces and
- a flexible agent platform allowing to distribute the components onto a network of computers, i.e., an information infrastructure, that has wireless connections to the nomadic user.

In Deep Map, artificial intelligence, is not used solely in a single component or as one monolith general problem solver. In contrast, AI appears as a necessary prerequisite in multiple components. In the Deep Map scenario, methods from AI, symbolic and sub-symbolic, are important for building smart components, that constitute the basis for the prestigious application presented here. In the following, we introduce several of these components.

We found that the tourism domain was extremely well-suited as a target for our prototype implementation, as we find nomadic and non-nomadic users (tourists), who are neither computer experts nor willing to be trained as such and have the need to access a variety of services from various domains both at home and while traveling.

2 The Deep Map System

2.1 An Intelligent Information Infrastructure

Deep Map as an intelligent information system that may assist the user in different situations and locations, has been developed with two faces:

- a web-based interface that can be accessed at home, work or any other networked PC

- and a mobile system that can be used everywhere else.

Both systems, however, are build on identical architectures and communicational protocols ensuring seamless information exchanges and hand-overs between the static and the mobile system. User profiles from the web-based system, for instance, should be also available for the mobile system and vice versa. The services provided to the user and the underlying knowledge bases of the system should have the same capabilities in each usage scenario. The main differences between the systems concern the interface paradigms employed and network- and performance-related aspects.

The Deep Map scenario is built upon a highly distributed, heterogenous and extremely dynamic environment. While in the desktop scenario, all components can be left on one computer, this might not be possible in a mobile scenario where only a small mobile device can provide the computational power for the basic components. Here, we have to be able to distribute the system so that heavy-weight components such as a full-grown geographical information system (GIS) resides on a server and the mobile device just carries a subset of the server information. In this case, the user can still access all data through wireless communication. Moreover, since wireless communication may be unstable, the system should not depend on mobile network connections and must have enough capabilities to self-organize accordingly.

In order to realize an intelligent infrastructure for this, we developed an agent platform that does not depend on a central management module. Deep Map utilizes an asynchronous distributed systems platform based on a distributed tuple space (Message Bus) paradigm, thus enabling the agents to join or leave during runtime.

In compliance with the Knowledge Sharing Initiative [6], agents can communicate via messages broadcasted through the Message Bus. The messages are structured in three layers:

- communicative layer, containing attributes such as sender or receiver,
- message layer, containing the performative or speech act [5], such as tell, ask, etc.
- content layer, containing the encapsulated information of the message.

This allows for a separation the message infrastructure from the message content and thus one sending agent does not have to know where the receiving agent of a message resides. Even more, the receiving agent that handles a message does not need to be fully specified. In our scenario, for instance, a request to a GIS can then be directed to the server GIS or to some small mobile GIS on the wearable computer depending on networking availability.

In order to allow agent communication on a higher semantic level, we specified a repository of hierarchically structured concepts, called Deep Map Objects (DMO). Agents can only construct messages whose content is composed of elements from this repository, i.e. the DMO represents our Ontology and Agent Interlingua.

2.2 The Agent Community of Deep Map

In general, Deep Map agents reside on three logical layers and are linked to each other through the agent platform described above:

- the interface layer (components that directly interact with the user such as the graphical or natural language interface),
- the cognitive layer (components that try to understand what the user meant and react accordingly, e.g. the presentation planer), and
- the service layer (components that provide basic services such as databases, geographic information systems and hotel reservation systems).

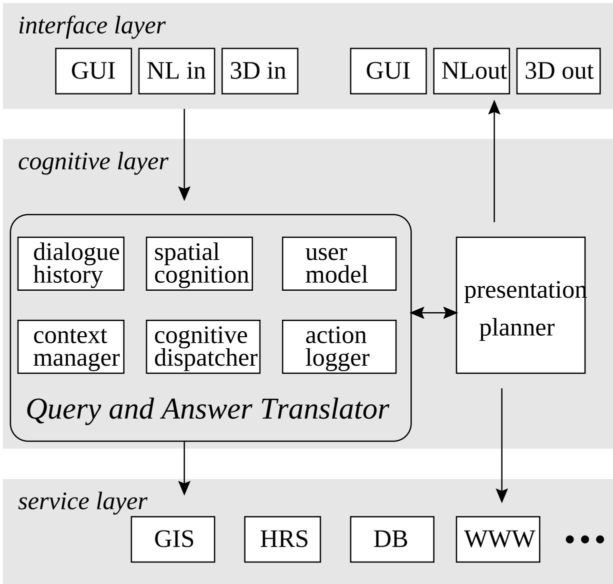

Figure 1. Agents in Deep Map are located in three layers: interface, cognitive, and service layer. The main direction of interaction goes from upper left (user input) through middle left (QUATRA components) to the service layer and back up on the right side to output to the user.

Figure 2.2 shows the three layer architecture of the whole system. The arrows depict the main directions of information flow from the user to the system, through the system and back to the user. An important set of agents resides in the QUATRA part, the query and answer translator, that aims at extracting the user's intention and translating the user's requests into system queries. These are then jointly solved by the system agents that provide the respective services, e.g. the geographic information system, historical and architectural databases etc. These responses are the, in turn, passed to the presentation planning module which is also notified about the nature and content of that response and generates an appropriate output for the user.

Note that only the main paths of communication are shown. Agents may mutually talk to every other agent. For instance the language output may need to ask back at the user models in QUATRA for information on the user. As described above, each instance may also be represented more than once in the same or in different variations.

It has already be mentioned that such an agent infrastructure needs an intelligent agent platform, but there is also the need for intelligence within the single agents. Depending on the tasks that have to be solved in each of them, they may use quite different AI methods to perform their tasks in an adaptive and autonomous way.

In the following, we describe some aspects of the user interface agents, cognitive and the service agents in more detail.[1]

[1] A full description of all agents would go far beyond the scope of this paper.

2.3 Smart User Interfaces for Every Use

The Web-based system focuses less on the use of natural language processing, even though it contains the same cognitive modules for processing written input. In many cases a "home" user in a *lean-forward mode* can gain faster access to the desired information by interacting with a graphical user interface, whereas a mobile user often prefers or even needs a hands-free approach to accessing information, and moreover may not desire to have his vision distracted (for example while driving a car or while walking). We focus on providing the Deep Map user with both a graphical and a natural language interface so that both is available and the appropriate modality can be selected according to the user's preference or her situation. On the first hand this enables the user to employ natural language as a means of interacting with the system, by simply uttering requests via spontaneous speech. On the other hand the system can make use of the natural language understanding capabilities of the user to convey the information computed. We see natural language as an advantageous bidirectional modality, albeit not exclusively. Depending both on the situation and on the content to be conveyed other modalities than speech can be preferable. In many cases the capability to generate multi-modal responses as well as the ability to process multi-modal input can be regarded to be an essential feature of a smart interface.

In a joint effort together with two cooperating institutes, the Interactive Systems Laboratory[2] and the DFKI[3] we have integrated components such as a spontaneous speech recognizer and syntactic and semantic parsing modules that produce language independent representations of user's utterances [7] that are further enriched by cognitive modules to be described below.

To be a able to render a system response in such ways as are suitable for a human recipient we have developed a natural language generation component that has the capability to create linguistic surface structures that identify, localize and describe objects and direct the user to desired locations. The main goal in implementing this generation module was to package information in such a way that it allows the user to create an appropriate mental model of the state of affairs to be conveyed in an efficient way. Looking at a sample input phrase such as "well, how do i get to that church", we can see that the system has to:

- be location aware, i.e. knows the user's location in the real world[4]
- be able to resolve different sorts of anaphora and references, based on the previous dialogue and actions of the user
- be able to access geographic information systems, e.g. to find a route to the desired location and suitable reference objects.

An additional prerequisite for creating an appropriate verbal response, is to be aware of the user's actual mental model and to search for referential entities that can safely be assumed to be part of the user's knowledge. In the example given above we use an estimation of the common ground both in selecting reference objects as well as for the linguistic realizations used in referring to these objects. The dialogue history and the user model, however, are not the sole sources of information for selecting appropriate reference objects further information is obtained by the cognitive components that trigger visibility and proximity analyzes in the geographic information system. Together these factors can are combined to compute individual degrees of applicability for the potential reference objects.

The natural language generation module is designed to adapt to the given richness or impoverishment of the input it is given, much in the same way as the graphical rendering components have to restrict their level of detail to bandwidth and display resources. Depending on the assortment of potential reference objects and the spatial relations that were to result from their employment, it uses a trade-off estimation for picking the most suitable reference object(s) in respect to the nature and quality of the ensuing spatial relation. This method guarantees the employment of *lesser* reference objects for the sake of a more expressive relation type, such as path relations [4] or the employment of less expressive relations for the sake of an extremely accessible reference object [2].

Additional adaption to the user and the communicative situation results in the generation of more or less elaborate confirmation and termination phrases [1]. At the moment we generate linguistic surface structures, in three languages - i.e. German, English and Japanese, based on the same language independent input. An additional feature is that we do not need to create Start-to-Finish instructions, but employ route segmentations performed by the cognitive modules. We can, thusly, guide the user in smaller steps towards the intended destination, thereby mixing graphic and linguistic information chunks. After the system has verbalized an instruction for a given route segment and displayed a map, the modules will wait until the user has reached the end of that segment, before it generates the next output, thereby keeping the memory load of the user minimal. If the presentation manager decides to make these instruction audible it notifies an articulation module that supervises the synthesis of these surface structures into speech. Otherwise the presentation manager can decide to have the instructions displayed along with maps, pictures and or animations.

2.4 Smart Spatial Services

So far, we demonstrated how AI methodology as intelligent agent technology, as cognitive modeling and for natural language processing can be used to build a powerful agent infrastructure and intelligent user interfaces. In the following, we briefly want to describe some aspects of applying intelligent methods in the service layer. Next to generating understandable spatial instructions, i.e. how to get from A to B, an important research goal within the project is to develop a system that is capable of generating individual tour proposals through a city based on the personal preferences and interests of a tourist. In order to achieve this, several problems have to be solved. These include the recognition of individual interests which is addressed in our efforts for building user models[5]. Building user models includes learning from observations and may employ a variety of learning techniques from symbolic to sub-symbolic methods. In order to get a working basis, however, we start by building stereotypes that reflect behavior and interests of tourists to Heidelberg. The statistical basis for these comes from a survey on about 350 international tourists to Heidelberg [3].

However, even if these interests are known there are several possibilities how to include them into a tour planning or tour proposing algorithm. First of all the range of possible attributes that may influence the choice for a particular section of a route have to be identified and

[2] ISL is at the University of Karlsruhe and at Carnegie Mellon University in Pittsburgh

[3] German Research Center for Artificial Intelligence, Saarbrücken

[4] Having a dGPS signal as the sole input for localizing the user at the moment, the system's position manager has to rely in inferences for estimating the likely orientation of the user in respect to the cardinal directions. For that, we employ vector information from previous positions and inferences based on the user's action, e.g. looking at specific objects or entering buildings etc.

[5] This work is done in a joint project with researchers from GMD-FIT, St. Augustin and will be described elsewhere.

modeled. Appropriate variables have to be included into the database and attached to the street network within the GIS. Such attributes include both ”hard” restrictions, or physically given attributes (like height, steepness, turn rules, legal rules, etc.) as well as a range of more dynamic and ”soft” parameters whose importance can vary extremely from person to person or time to time. Such parameters could include esthetic aspects, the social milieu of the area, dislike of motorized traffic or preferences for areas with high degree of architecturally interesting buildings or just a high rate of scenic viewpoints. If these parameters are included in the tour planning algorithm the task to propose a whole tour is still twofold. It is not only necessary to find the most appropriate (as opposed to merely the fastest, as it is realized in conventional tour planning software) route from location A to B, but also to select the possible stops (sights, restaurants, etc.) to visit, as well as the number of them, according to the tourists interests. Right now we have developed two algorithms, which take into account a range of hard and soft parameters for each street section. One is based on interpolating so called service areas on the street network, the second is based on buffering the street network. Within these, techniques optimization methods such as simulated annealing and genetic/evolutionary algorithms may be employed for finding good solutions for these NP-complete tour planning problems.

Figure 2. Calculated tours for a pedestrian. Both tours cover the same places (filled circles), but differ in preference settings. The upper tour tries to avoid noise and pollution and lower tour prefers attractive areas.

Figure 2.4 shows two tours computed by Deep Map. In both tours, the same spatial objects are visited, but the soft parameters are weighted differently. In the one tour, the preference is triggered such that the avoidance of noise and pollution leads to the selection of streets with less traffic and in the second example has a higher weight on visiting interesting sites. Other soft parameters can also have an effect of the selection of the objects that are considered for being visited. Thematic interests such as interest in particular architectural poques, are a good example that may be used for proposing individual tours according to the user’s interest. Another personal choice is the means of transportation (car, foot, bike and wheelchair) that can be chosen individually and is considered in selecting appropriate paths.

2.5 Other Applications of AI in Deep Map

We can only highlight a few components in Deep Map where AI plays a role. Tour planning and intelligent user interfaces are just some aspects of smart services where multiple AI techniques come into play. There are many other aspects in the GIS services that incorporate *fuzzy* knowledge, reasoning and intelligent heuristics. For instance the generation of maps is a well known field where AI methods have been applied. In Deep Map, the task is to render maps of specified areas such that it visualizes the information that is best suitable for reaching a communicative goal. Therefore, decisions are needed for which data has to be shown explicitly and which have to be highlighted.

But also (non-spatial) databases require AI techniques in order to provide smart services for users that want to access complex knowledge that is distributed over heterogenous data sources. For historical data temporal reasoning comes into play, access to multiple databases requires shared ontologies, and access to data sources provided by others require intelligent information integration methods.

3 Deep Map: Challenges and Future Directions

The Deep Map system as described above represents a research framework that is under continuous further development. The parts presented here represent one segment of the work that has been done at the European Media Laboratory. Some components such as those for NLP input, spatial reasoning and user modeling are done in collaboration with or contributed by other institutes. Deep Map addresses some very hard ”AI-complete” problems such as natural language understanding and we are aware of the fact that some of our for reaching goals cannot be solved within a short time period. The aim of Deep Map is to address these research issues and to integrate the single steps in the direction leading towards a complete system. This methods helps to grasp the difficulties of a creating real world application and also its interim results serve as a demonstrator for the each single component.

So far, we developed a running system that can be used as a Web-based system for tour planning and as as a mobile navigation and information system. The mobile system shown in Figure 3 uses a wearable PC as hardware platform with a LCD display and a headset plus microphone. Because of the flexibility of our architecture and agent system, we can easily port the whole system to other hardware configurations.

In its current state, Deep Map, can understand and speak English. First components for for German and Japanese as additional languages have been developed. The vocabulary and the demo scenario is the castle area and old town of Heidelberg.

Even though the emphasis of this paper lies on the application of AI in such a system as Deep Map, it should be noted that a great deal of effort had to be invested into the integration of all its components. It is, in fact, a non-trivial task to integrate a GIS that is based on a commercial NT platform with a Linux-based speech-recognizer, some quite complex reasoning components and other modules on a mobile platform without loosing too much performance. This task of integration and performance optimization will also be a relevant challenge for future developments and is integral part of applying AI systems and components within real world applications.

But also the single components will benefit from further extentions. We are confident that speaker-independent spontaneous speech recognition will be solved with this decade. Natural language understanding, however, will remain an long-term research issue.

Figure 3. The Deep Map system in mobile use. The Tourist has a head set for natural language interface, a display on his arm and a wearable computer mounted on his belt (Xybernaut MAIV pentium system).

3.1 Conclusions

In this paper, we gave an overview of the Deep Map system. Deep Map is a very ambitious project where we want to build a visionary assistance system. This vision incorporates information infrastructures as postulated by the new paradigm of the *disappearing computer*. This new paradigm of human computer interaction covers seamless information access everywhere, user and situation adapted services as well as intelligent user interfaces. Artificial intelligence is a necessary component for many components of such a system. There is, however, not one single AI method that dominates in the overall application. In contrast, an array of AI methods appear in the single components.

The infrastructure, for instance, is based on the paradigm of intelligent agents, where software components are conceived as autonomous entities that collaborate in order to achieve a goal and communicate using shared ontologies. Natural language generation uses cognitive and psycholinguistic findings in order to produce linguistic output that is appropriately tailored for the human user. Next to cognitive issues such as natural language understanding or presentation planning, systems residing on the so-called *service layer* also need AI methods for information retrieval, personalized services and tour planning methods. Together the whole ensemble of intelligent components, methods and architectural features sets the stage for novel information systems that can be perceived by human users as useful extensions of their own resources rather than technical implements.

Acknowledgments

The research presented here has been funded by the Klaus Tschira Foundation and the bmb+f in the framework of the SmartKom, EMBASSI and Deep Map projects.

REFERENCES

[1] J. Bourne, ‘Generting effective instructions: knowing when to stop’, *Phd Thesis, Computer and information Science ept. University of Pennsylvania*, (1998).
[2] K.-P. Gapp, ‘Basic meanings of spatial relations: Computation and evaluation in 3d space’, in *Proceedings of AAAI-94*, Seattle, WA, (1994).
[3] E. Kaul, ‘Sozioökonomische Kategorisierung von Heidelberg Touristen’, *Diplomarbeit. Geographisches Institut der Universtät Heidelberg*, (1999).
[4] C. Kray and A. Blocher, ‘Modeling the basic meanings of path relations’, in *Proceedings of the 16th IJCAI*, pp. 384–389, San Francisco, CA, (1999). Morgan Kaufmann.
[5] John R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, 1969.
[6] External Interfaces Working Group The DARPA Knowledge Sharing Initiative. Specification of the kqml agent communication language, 1993.
[7] M. Woszczyna, ‘Fast speaker independent large vocabulary continuous speech recognition’, *Ph.D. Thesis, University of Karlsruhe, Germany*, (1997).