The Pragmatic Web: Agent-Based Multimodal Web Interaction with no Browser in Sight

Alexander Repenning & James Sullivan

University of Colorado, Computer Science Department, Boulder, USA

ralex@cs.colorado.edu

Abstract: To a large degree information has become *accessible* – anytime, anywhere – but not necessarily *useful*. Unless the right information is presented at the right time, in the right place, and the right way information may simply become unwelcome noise. The Pragmatic Web is about giving control to information consumers so they can customize *how to use* information. The original Web – The Syntactic Web – has given most of the control to the information producers but, aside of trivial font and plug-in options, has provided little control to information consumers. The Semantic Web better separates content from presentation but has not yet received general acceptance and lacks end-user customization tools. The Pragmatic Web – presented here – is about providing control to information consuming end-users by enabling them to express computationally how to turn *existing information* into *personally relevant information*.

Keywords: auditory input and output, context aware computing, end-user programmable agents, end-user development and adaptation, sensors and actuators, virtual reality and 3D interfaces.

1 Introduction

Information is no longer a scarce resource. However, this achievement is largely useless if information is not provided in a format tailored to the user. Most people in the United States, for instance, now have access to online information thanks to inexpensive computers and information appliances (Fox et al., 2000), and public access to computing resources in libraries and other institutions. Web-based information can be accessed – through wires or wirelessly – from desktop computers, laptops, PDAs, cell phones, and specialized information appliances. However, *ubiquitous information access does not imply universal information access*. The representational formats chosen by information producers often do not match the needs of information consumers. For instance, a Web page may contain highly relevant information to a user, but this information cannot be accessed if the user cannot see or cannot read. Reasons for producer/consumer *information representation mismatch* include:

- **Wrong Modality**: Blind users cannot read textual descriptions. Automatic text-to-speech interfaces may be able to verbally convey the textual contents of a Web page to users, but if the Web page is formatted for visual access, the sequential presentation of information as speech may be unintelligible or inefficient.
- **Wrong Language**: Crucial explanatory text may be provided in the wrong language. Less than 15% of U.S. Web sites contain Spanish translations (Lamb, 2001).
- **Wrong Nomenclature**: Information may be expressed in an unfamiliar measurement system. The translation of Celsius to Fahrenheit or kilometers to miles, while scientifically trivial, may present a serious impediment to many users.
- **Wrong Time**: Information may be correct, relevant and readable, but presented at the wrong time. Stock information, for instance, is most useful when presented in real time.
- **Wrong Format**: Information can look great on a large computer monitor, but be completely unsuitable for small information devices such as PDAs and cell phones.

A mismatch between information presentation and the formats required by an information consumer can be difficult to address with a traditional Web browser. For economic reasons, a producer may choose to use a single representation scheme that addresses only the needs of an anticipated majority of information consumers. Because creating and maintaining multilingual Web sites can be costly,

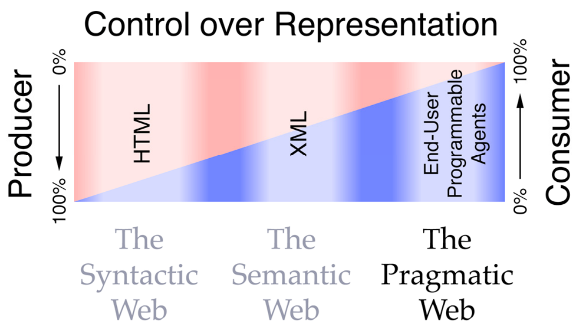

Figure 1: Distribution of Control over Representation shared by Information Producer and Information Consumer.

information consumers who are not proficient English readers have few options with this model.

In the case of a person with a cognitive disability planning to use the public transportation system, there is a good chance that essential information exists on the Web but cannot be accessed meaningfully with conventional information technology. The goal of this proposal is to extend control to information consumers and let them use information in fundamentally new ways that might not be anticipated by information producers.

The ultimate questions involve *who* controls information representation and *how* the information is processed. The Control over Representation diagram (Figure 1) illustrates a continuum of control that ranges between two extreme positions. From left to right, it identifies conceptual as well as technical frameworks in order of increasing information consumer control: the Syntactic Web, the Semantic Web and the Pragmatic Web.

The Syntactic Web. In this first generation of Web technology, a markup language (HTML) is used to define content at a high level of detail. This syntactic level controls the appearance of information. Information producers define content, font selection, layout, and colors. Information consumers have limited control over representations in their browser, including adjusting the size of fonts, and enabling/disabling animations and plug-ins.

The Semantic Web. According to Tim Berners-Lee, the Semantic Web (W3C, 2001, Berners-Lee et al., 2001, Berners-Lee, 1999) will "radically change the nature of the Web." (Berners-Lee et al., 2001) The formal nature of representation languages such as the eXtensible Markup Language (XML) and the Resource Description Framework (RDF) make Web-based information readable not only to humans, but also to computers. For instance, semantic-enabled search agents will be able to collect machine-readable data from diverse sources, process it and infer new facts. Other research projects, such as the Avanti project, have studied how to separate web content from display modality to better serve the sensory and perceptive abilities needs of users (Stephanidis and Savidis, 2001).

Unfortunately, the full benefits of the Semantic Web may be years away and will be reached only when a critical mass of semantic information is available. Critics of the Semantic Web (Frauenfelder, 2001) point out the enormous undertaking of creating the necessary standardized information ontologies to make information universally processable.

The Pragmatic Web. In contrast to the Syntactic and Semantic Web the Pragmatic Web is not about form or meaning of information but about **how information is used**.

The Pragmatic Web's mission is to provide information consumers with computational *agents* to transform *existing information* into *relevant information of practical consequences*. This transformation may be as simple as extracting a number out of a table from a single Web page or may be as complex as intelligently fusing the information from many different Web pages into new aggregated representations.

This agent-based transformation needs to be extremely flexible to deal with a variety of contexts and user requirements. An agent running on a desktop computer with a large display may utilize rich graphical representation versus an agent running on a cell phone with a small display may have to resort to synthesized text information to convey the same information.

The Pragmatic Web research explores the practice of using information and the design of tools supporting this process.

Instead of the traditional "click the link" browser-based interfaces, agents capable of multimodal communication will provide access to Web-based information. Agent communication methods include facial animation, speech synthesis, and speech recognition and understanding. End-users or caretakers will instruct agents to transform information in highly customized ways. Agents will work together to combine information from multiple web pages, access information autonomously or triggered by voice commands, and represent synthesized information through multimodal channels.

End-User Customization (Nardi, 1993, Jones, 1995) will be an integral part of the Pragmatic Web. Successful development of end-user customization will make giant steps toward an Every Citizen

Interface (ECI) (Committee et al., 1997) by letting minority groups of information consumers, possibly down to the single individual level, obtain ways to control information representations in accordance with their specific needs. This computer-supported *Information Processing* is a form of knowledge management (Sumner, 1999, Murray, 1999) that turns raw data into information. End-user customization will let users specify *where* information is accessed (e.g. part of an existing Web page), *how* it is accessed (e.g., voice activated), and how information is further *processed*. For instance, a Pragmatic Web application could run on a wireless PDA equipped with GPS to help a person with a cognitive disability navigate through town using the local public transportation system.

The Pragmatic Web does not intend to subsume the Syntactic Web or the Semantic Web. On the contrary, the Pragmatic Web will initially work with the Syntactic Web by letting end-user customizable agents extract information out of existing (HTML) Web pages (Inmon, 1996). When the Semantic Web reaches a minimal critical mass, the Pragmatic Web will then utilize the Semantic Web with agents that access ontologies and make inferences based on these representations.

This article describes how end-user programmable agents allow users to change modalities to make information show up at the right time, and to fuse information from multiple sources into new formats. To illustrate the Pragmatic Web framework, two example applications are presented. The Mountain Bike Advisor keeps track of current weather conditions and personal biking preferences. A voice interface recognises requests, sends out agents to access remote sensor information and applies user-defined rules to interpret information pragmatically. The Boulder Transportation system tracks GPS-equipped busses in real time, renders a 3D visualization, and interprets bus location information to make navigational recommendations for persons with cognitive disabilities.

2 Example Applications

Two early Pragmatic Web applications are presented here to illustrate how agents are being programmed by end-users, how these agents access information in existing Web pages, how they process that information, and finally how the interact with the users by interpreting the information found. Both applications are working prototypes.

2.1 Boulder Mountain Bike Advisor

This AgentSheets-based application connects real-time Web information with speech recognition. A user asks, "Where should I go mountain biking?" Several agents located on a map of Boulder County react to this voice command (Figure 2). These agents are representing locations that are possible candidates for biking and also feature real time, Web accessible weather information sensors.

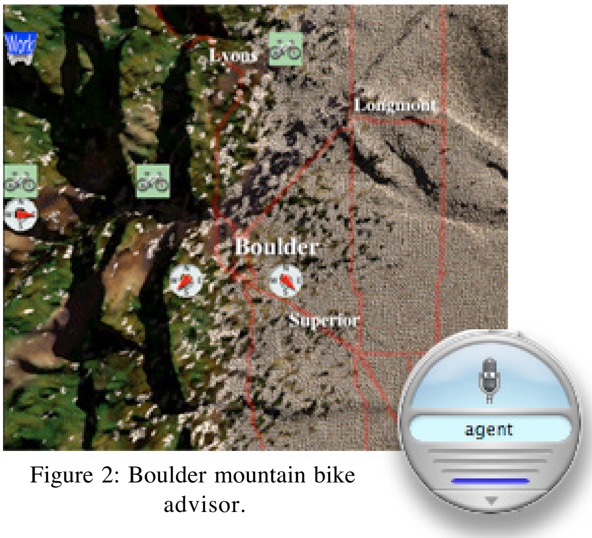

Figure 2: Boulder mountain bike advisor.

Rules, previously defined by the users, capture pragmatic interpretations. For instance, an agent may reply (using speech output): "It's really nice up here at Betasso but you should bring a jacket because it's a little windy."

Behind the Scenes

How does all of this work? The information used by the Bike Advisor originates in a number of Weather stations featuring a large array of sensors accessible via Web pages. The network of weather stations in Boulder County, Colorado is relatively dense so that for most bike trail locations a weather station can be found to sufficiently well estimate current weather conditions. The C1 Niwot Ridge Weather station is closest to the Sourdough Bike Trail and features a well-organized Web page (Figure 3).

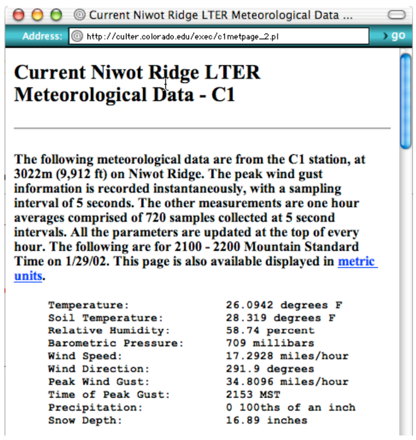

Figure 3: NCAR's Foothill Weather Sensor Web Page serves as input for agents to make recommendations.

Our scenario begins with the user speaking the word “Biking.” All the agents, represented as icons on the map (Figure 3), listen to voice commands. Agents in AgentSheets are programmed by end-users in Visual AgenTalk (Repenning and Ambach, 1996). Visual AgenTalk is a rule-based language. The first rule of the Sourdough biking advisor agent has a speech recognition condition becoming true as the result of what the user just said. The agent now triggers a second rule group called “check” which includes two conditions accessing the C1 weather station Web page to extract the current temperature and wind speed information. The Fahrenheit information is numerically converted into Celsius information and, using text to speech, announced to the user: “Temperature at Sourdough is currently –3.2 degrees Celsius.”

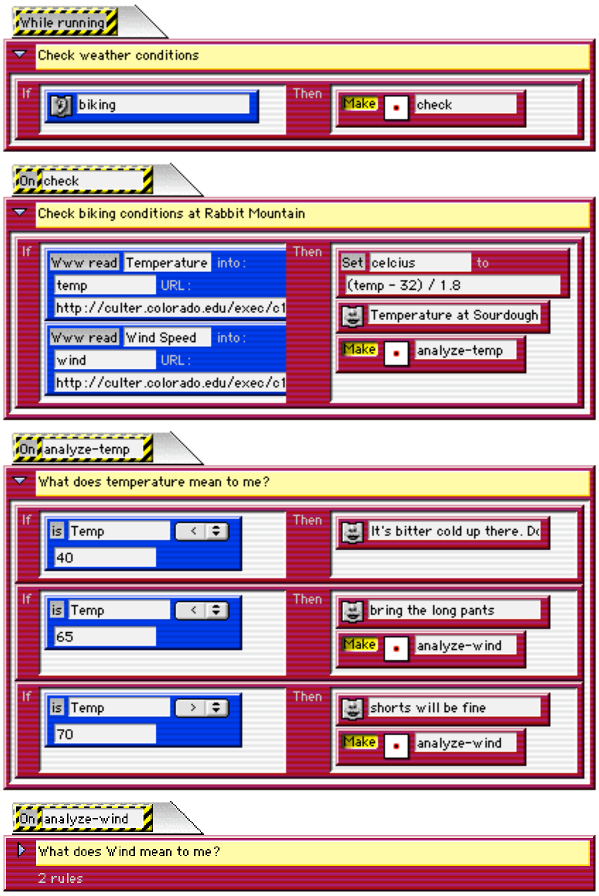

Figure 4: AgentSheets end-user programmable authoring environment illustrating if-then rules.

Temperature and wind speed are further interpreted by calling a third rule group. This is the pragmatic part of the interpretation using temperature and wind thresholds that are only relevant to the user who has expressed these rules. Unlike the objective part of the rule which merely communicated the numerical value of the temperature, the pragmatic part is directly employed to reach a decision. In our case since the temperature (in Fahrenheit) is less than 40 degrees the agent advises against a bike ride at Sourdough: “It's bitter cold up there. Don't come here!” In more moderate cases the agent would have recommended to bring additional clothing such as wind stoppers in case the wind speed exceeds a different threshold.

There are two important things to note here. First, the information returned by the agent is of a highly pragmatic nature that is directly relevant to the user and to the goal of mountain biking. These rules are easily created and modified by the user using a highly visual programming environment (Figure 4).

Second, the interface, input and output, can be tailored to suit user needs and delivered in multi-modal forms, including visual and speech. For example, the map display while helpful to locate a specific trail, is not necessary to interact with the system. Indeed the shift in modality from text to speech allows the entire interaction to take place over a cell phone without the need for any display or a traditional browser, nor any need to modify the original Web page in any way.

2.2 Mobility Agents

Navigating through a city public transportation system can be a daunting challenge for a person with memory and attention problems due to cognitive disabilities. The Mobility-for-All research project sponsored by the Coleman Institute for Cognitive Disabilities is studying how personalized mobile information technologies can assist such travelers by eliminating information overload from traditional navigational artifacts (Sullivan et al., 2002).

artifact	Purpose
maps	spatial relationships between one’s current location and destination; identify routing options; provide an abstract means to assess overall trip progress.
schedules	temporal information about route availability at a given day and time.
landmarks	to confirm global progress and anticipate important events or tasks that will come next, such as prepare to get off, etc.
labels and signs	to understand the local environment, including: current location, where to meet transportation vehicles; identify the “right” vehicle; where to get on and off; where to pay; etc.
clocks	to synchronize schedules with physical events, including transportation vehicle arrivals and departures.

Table 1: Essential navigation artifacts commonly found in public transportation systems

As buses travel on Boulder Colorado city streets, they report their Global Positioning System (GPS) locations on a wireless network. Agents track bus locations, and use this information to generate personally relevant "just-in-time" attention and memory prompts in a visual and/or auditory form that can be tailored to the mobile user's needs, abilities, and preferences.

Architecture

Our current prototype consists of a number of connected components (Figure 5). The existing infrastructure includes the GPS-equipped transportation system of 27 busses, transceivers that connect the GPS sensors (one per bus) with a central GPS information receiver, and a Web server that puts the positional information onto the Web. Every two seconds each bus sends an update of GPS position, heading, identity and speed. Currently, this information is gathered on an operations console where it is both archived and visualized as dots moving on a map. These dots are observed by a human dispatcher, who informs drivers of bunching and other problems. To a caretaker or our traveler, a person with a cognitive disability, this representation of information (at the syntactic level) is without meaning or use.

Mobility Agents

The Mobility Agent server is the central architectural component that mediates all communications. Mobility Agents running on this server can read the GPS information (bus ID and location) from the provider server and track buses currently in the transportation system. The server in turn feeds real time bus and traveler information to the 3D Transportation Situation Viewer.

When a traveler makes a choice, such as picking a destination, the server receives that choice along with the traveler's location. Mobility agents generate appropriate responses, which the server then transmits to the traveler via events that contain multimodal instructions, reminders, or prompts. The server can also deal with traveler confirmations and panic alarms.

The responses and information provided to the traveler vary with the user profile, and depend on mobility agents with custom settings for that specific traveler. User profiles for multiple travelers are maintained on the server and contain personalized information such as schedules, typical itineraries and destinations, and contact information. The behavior of mobility agents can be readily modified, so the server also provides caregivers customization capabilities. The current version of the Mobility agents uses Visual AgenTalk rules similar to the ones shown previously to map data at the syntactic level (the GPS location strings) to pragmatic level (e.g., "your bus is here").

Mobility agents produce signals that: 1) define events available to the traveler; 2) create responses in different modalities; and 3) provide exception handling and error resolution mechanisms. A caregiver then personalizes the mobility agents for particular travelers by selecting settings appropriate to travelers' needs and abilities.

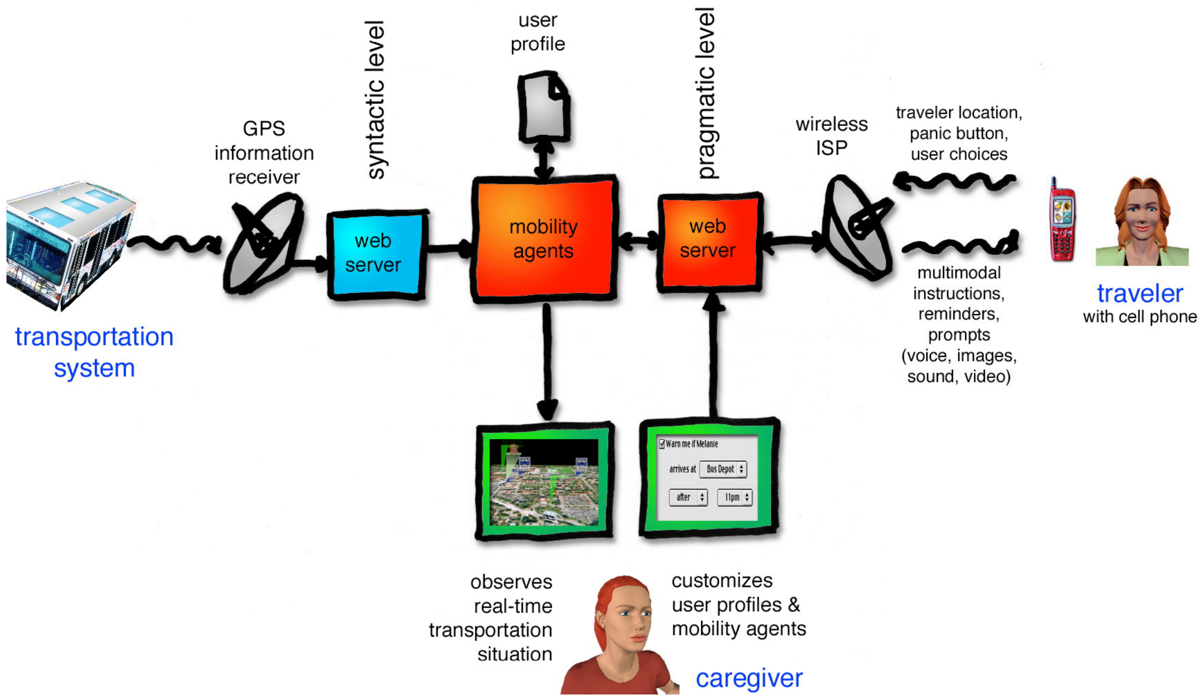

Figure 5: Mobility Agent Architecture

Real-Time Transportation Viewer

Caretakers see a real-time visualization of the transportation state via the Real-Time Transportation Situation Viewer. The client tracks the buses in the transportation system as they run their routes. The buses are depicted as moving 3D objects on a map rendered in OpenGL. Bus stops are clearly marked with superimposed bus stop signs. The travelers' locations are tracked and represented on the 3D map in real-time. The Mobility Agent server provides bus and traveler location information.

Figure 6: A mobile agent-based prototype that provides just-in-time prompts for people with cognitive disabilities as they use public transportation systems.

Transportation Visualization lets caregivers monitor travelers, assess a traveler's ability to effectively use the transportation system, detect difficulties in daily routines, and receive emergency notifications. Viewers can assume different perspectives in the 3D word including fixed camera positions (e.g., at the a bus stop), and cameras tracking objects (e.g., bus driver, or traffic helicopter perspective).

Mobility Agent Customization Client

The Mobility Agent Customization Client would let caregivers customize user profiles and mobility agents through a browser interface. User profiles can contain personal information such as typical itineraries; such information can be used to detect deviations from a normal route and evoke error-handling procedures. The Customization client would provide an interface for customizing the behavior of mobility agents to map specific situations to appropriate actions. The client would allow the specification of normal as well as exceptional responses to traveler actions and choices. For example, in a normal situation when everything goes according to plan, the caregiver can define the sequence of actions and prompts to be given to Melanie, the cognitive disabled teen, after she chooses her destination (home, in the case illustrated by the scenario). This might include: 1) finding the location of the right bus; 2) informing Melanie how close it is to the bus stop she is at; 3) warning her when it is arriving; 4) guiding her to board the bus; 5) informing her when her destination is approaching; and 6) reminding her to gather her belongings before leaving the bus. In an exception-handling situation, such as Melanie boarding the wrong bus, the caregiver can customize increasingly intrusive prompts. These might include: first warning her that she is on the wrong bus with subtle sound notifications; then advising her to talk to the driver; and then, if the situation is not resolved, having the client application on the cell-phone send a panic alarm to the caregiver. The caregiver could then contact the traveler directly by phone. Events and reactions are tailored to travelers with different cognitive abilities. For example, an individual with the ability to follow more complex instructions might be advised to get off the bus, cross the street, and catch a bus going in the other direction.

Wireless device

The client application running on a wireless device accepts a traveler's input choices and provides instructions, reminders and prompts to the traveler using multimodal means of communication, such as voice input and output, sound, images, and movies.

Our current prototype is based on a cell phone emulator running on a laptop equipped with a wireless network. The prototype is functional in the sense that it is connected to the Mobility agents, receives information based on the actual bus locations and allows users to define their goals.

The prototype illustrates agent-based components of a mobile architecture (Figure 7). The person simulated in this prototype (Melanie) is assumed to be a teen with developmental disabilities including attention and memory deficits. Melanie can be "directed" to a bus stop where a bus is approaching, and the prompting sequence is "triggered" by selecting a destination option on her phone.

As the simulation runs, Melanie's mobile phone generates visual and auditory prompts triggered by real world events. Prompts are generated to "get ready" for her approaching bus, "please board now" when the bus stops at her location, "please pull the stop cord and prepare to get off" as the bus approaches the destination stop, "please get off here" at the destination stop, and finally, "don't forget your backpack." As Melanie performs these tasks, she is also rewarded with reinforcing praise.

The 3D visualization component could also provide bus system operators or waiting passengers with an overview of bus locations and status. Observers can watch real time traffic, play back recorded data or assume different camera perspectives

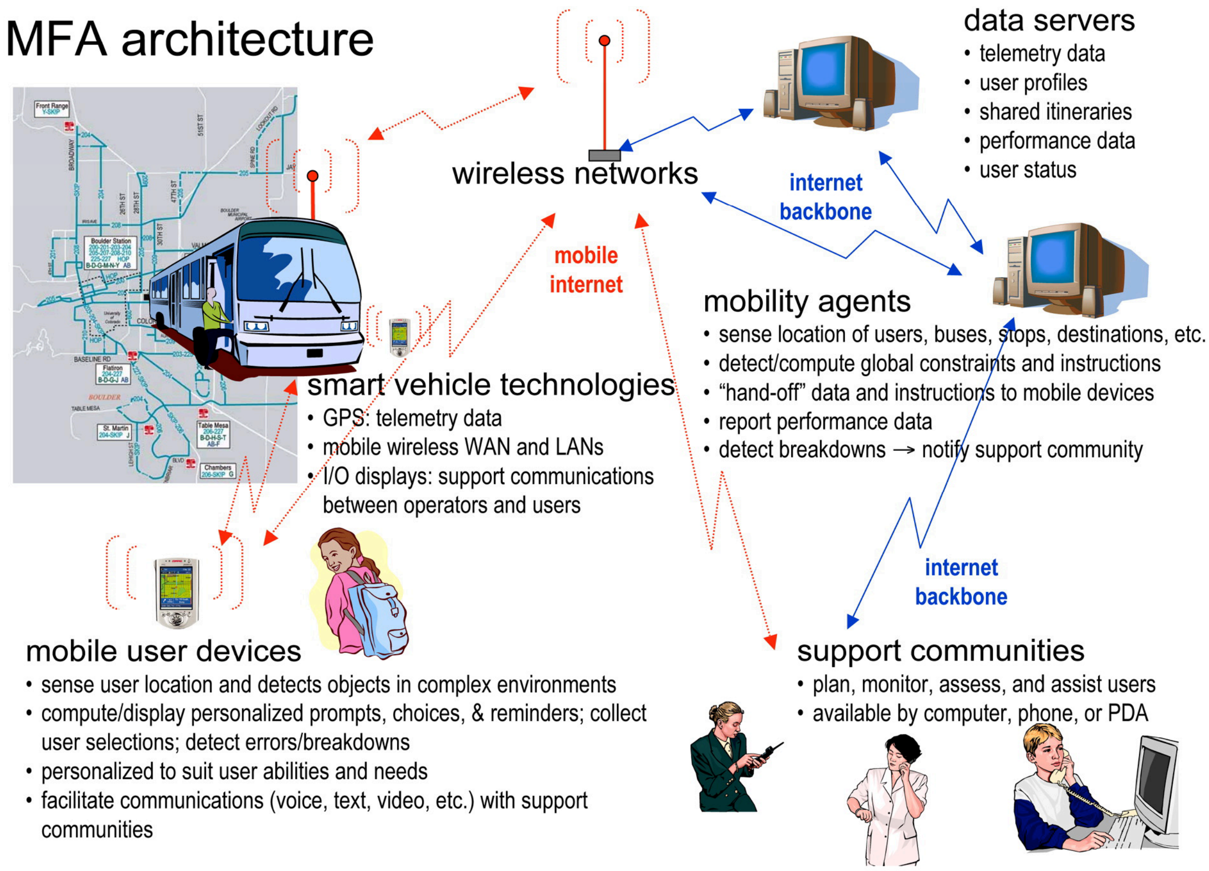

Figure 7: A mobile architecture for locating and delivering traveler information – and finding help if something goes wrong.

(e.g., birds-eye, bus stop perspective, bus driver perspective). Bus users can locate relevant busses based on their current position and bus identification information. End-user development tools allows care givers or service-providers, to specify rules that turn the general bus information space into personally relevant, pragmatic information communicated through cell phones.

3 Discussion

A big concern with parsing general Web pages is that the location or the format of a Web page may change. Indeed this is a valid problem, which, depending on the severity of the change and the robustness of the parsing approach used, may result in information that no longer can be accessed, or, worse, may result in reporting the wrong information. One way to think about the problem is to assume that changes to Web information which would confuse agents are also likely to confuse and even upset real users. For instance, these days Web designer are more careful with the creation of URLs since they know that people share them in emails and collect bookmarks. A Web site design strategy that frequently changes the location of essential Web pages is extremely likely to upset users.

A different approach to make parsing of Web pages more robust is to factor out Web page depending parts, which can be maintained separately. AgentSheets agents can be freely distributed through the Behavior Exchange (Repenning and Ambach, 1997). This way a community of service providers can take over the responsibility of centrally maintaining Web page depending agents. End-user developers of agent-based applications may choose to download the latest version of agents just in time. In AgentSheets, for example, this is simple since agents can download other agents.

The Semantic Web with its formalized interface could provide a more robust approach. The Pragmatic Web framework can and should use the Semantic Web but at least for now may need to fall back in most cases onto the Syntactic Web and more informal parsing simply because the majority of Web information is not yet available in semantic form.

4 Conclusions

The goal of The Pragmatic Web is to provide more control to end-users with respect to using information. Relatively simple end-user programming techniques allow end-users to transform general information available on the Web into personally relevant information accessible via wireless devices. End-user control - where to access information, how to process information, how to invoke information access, when to present information, how to process information and how to present information.

We have built a number of applications that, using relatively simple end-user programs, have wrapped up existing Web information in radically different interfaces. The resulting shift is not about a quantitative information access improvement but about a qualitative shift in affordances. The applications built are encouraging but should just be considered simple instances of the much larger Pragmatic Web framework.

5 Acknowledgements

This research has been supported by the National Science Foundation (EIA 0205625, DMI 0233028) and by the Coleman Initiative.

6 References

Berners-Lee, T. (1999) Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor, Harper, San Francisco, CA.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web, *Scientific American*.

Commission on Physical Sciences. (1997) *More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure,* National Academy Press, Washington, D.C.

Fox, A., Johanson, B., Hanrahan, P. and Winograd, T. (2000) Integrating Information Appliances into an Interactive Workspace , *IEEE Computer Graphics and Applications,* **30,** 54-65.

Frauenfelder, M. (2001) A Smarter Web, *Technology Review*.

Inmon, W. H. (1996) The Data Warehouse and Data Mining, *Communications of the ACM,* **39,** 49-50.

Jones, C. (1995) End-User Programming, *IEEE Computer,* **28,** 68-70.

Lamb, E. (2001) Web content struggles to go worldwide , In *Red Herring*, Vol. 91, pp. 38-39.

Murray, A. J. (1999) Knowledge management and consciousness, *Advances in Mind–Body Medicine,* **16,** 233-237.

Nardi, B. (1993) *A Small Matter of Programming,* MIT Press, Cambridge, MA.

Repenning, A. and Ambach, J. (1996) Tactile Programming: A Unified Manipulation Paradigm Supporting Program Comprehension, Composition and Sharing, *Proceedings of the 1996 IEEE Symposium of Visual Languages*, Boulder, Colorado, pp. 102-109

Repenning, A. and Ambach, J. (1997) The Agentsheets Behavior Exchange: Supporting Social Behavior Processing, In *CHI 97, Conference on Human Factors in Computing Systems, Extended Abstracts*ACM Press, Atlanta, Georgia, pp. 26-27.

Stephanidis, C. and Savidis, A. (2001) Universal Access in the Information Society: Methods, Tools, and Interaction Technologies , *International Journal of Universal Access in the Information Society,* **1,** 40-55.

Sullivan, J., Fischer, G., Binder, T. and Gregory, J. (2002) Human-Centered Public Transportation Systems for Persons with Cognitive Disabilities - Challenges and Insights for Participatory Design, In *7th Participatory Design Conference* (Ed, Wagner, I.) Malmo, Sweden, pp. 194-198.

Sumner, M. (1999) Proceedings of the 1999 ACM SIGCPR conference on Computer personnel research, In *Proceedings of the 1999 ACM SIGCPR conference on Computer personnel research* New Orleans, LA USA.

W3C (2001), http://www.w3.org/