Harnessing Language in Mobile Environments

Boris Katz, Gary Borchardt, Sue Felshin and Federico Mora
MIT Computer Science and Artificial Intelligence Laboratory
{boris, borchardt, sfelshin, federico}@csail.mit.edu

Abstract

We describe StartMobile, a prototype system that enables users of cellular telephones and other mobile devices to access information, create information and execute commands on the basis of written requests expressed in natural language. Interacting with mobile devices through the use of language offers several potential benefits that include: scalability in handling huge numbers of concepts and operations, substantial device-independence, and opportunities for remote interaction. StartMobile uses the START information access system to perform initial interpretation of requests and to generate responses to general information requests. Requests that require further interpretation and/or fulfillment in the distributed mobile environment are encoded in a new, language-based intermediate representation called Moebius, then conveyed to appropriate devices and systems, where special-purpose software serves to complete the interpretation and fulfillment of those requests.

1. Introduction

Users of cellular telephones and other mobile devices are faced with an ever-increasing range of capabilities at their disposal, and this development will soon overwhelm traditional user interfaces. Indeed, it is already the case that many operations available at the programming level within mobile devices are not accessible through devices' user interfaces. New mechanisms are required that enable mobile users to access information, create information and perform a full spectrum of actions with ease. We feel that a language-based interface is an ideal candidate for this task. First, language is well-suited to humans: it is expressive, compact and easy to use. In particular, language enables humans to articulate requests without having full knowledge of the range of capabilities offered by a system and without knowledge of specific terminology used to name those capabilities. Second, it is also within the capability of machines to handle requests posed in language, using emerging information access technology and related elements of language processing technology. Third, language-based interaction offers additional benefits of relative device-independence and support for remote interaction.

We have implemented a system called StartMobile that handles a range of language-based requests in mobile environments. StartMobile uses the START information access system [1][2] as a first stage in the processing of user requests. START performs an initial interpretation of the requests, and if these requests concern the retrieval of general information from the World Wide Web or other sources, START obtains the information for presentation to the user. If it is not possible to complete the interpretation of the requests, however, or if the requests involve actions that must be performed on the user's mobile device, START encodes the user's requests in a language called Moebius, which has been designed to convey natural language requests in various stages of interpretation between systems and devices. Finally, software that resides on the user's mobile device completes the interpretation of user requests, if necessary, and performs required actions to fulfill those requests.

The StartMobile system currently supports a range of activities on several models of mobile phones:

- retrieving general-purpose information for the benefit of the user,
- retrieving contact and calendar information stored on the user's mobile device,
- retrieving text messages and managing the user's text message inbox,
- placing phone calls from the user's mobile device,
- creating reminders on the user's mobile device or on other users' mobile devices,
- taking pictures with the mobile device's camera,

- modifying device settings,
- accessing position information and displaying associated map and direction information on the user's mobile device, and
- retrieving video tutorials for presentation to the mobile user.

Figure 1 illustrates the StartMobile system in action. (The figures in this paper were created using a screen capture mechanism to record display interactions on operating mobile phones.) Here, the mobile user has asked the system to list the user's contacts at a specific company. StartMobile performs the requested operation and displays the results to the user.
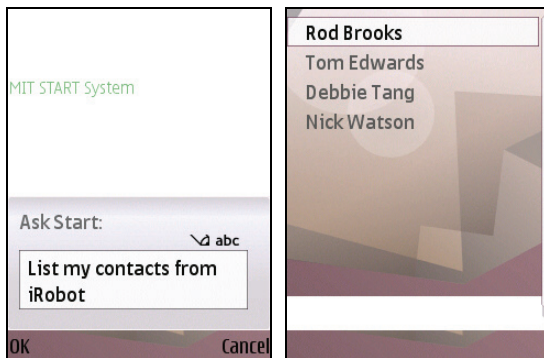

Figure 1. StartMobile responding to a request by performing a search within the contacts database on the user's mobile device

StartMobile allows its users to enter requests in a range of variant forms. For the example illustrated in Figure 1, some of these variant forms are:

Who works for iRobot?
Who do I know at iRobot?
Which of my friends work at iRobot?
Show my colleagues from iRobot.
...

Section 2 of this paper provides background information on the START information access system. Next, section 3 describes the use of START within StartMobile to retrieve general purpose information and, in combination with software on the user's mobile device, retrieve information from the device and perform actions on that device. Section 4 describes the Moebius representation language and StartMobile's use of Moebius in the processing of ambiguous requests. Section 5 discusses alternate configurations of systems and devices supported by the StartMobile design, plus additional types of capabilities supported by this design. Sections 6 and 7 then present continuing research and contributions of the work.

2. The START system

Providing a foundation for this work is MIT's START information access system. START has been in development for approximately 20 years and, in its most general application, is available as a public server on the World Wide Web at http://start.csail.mit.edu/. For the StartMobile application, START has been adapted to communicate with mobile devices, both in terms of connectivity and in terms of output presentation for small screens; in addition, START's traditional functionality of providing language-driven access to information has been extended to the execution of actions in response to requests.

In its traditional role, START accepts English questions and offers responses that draw upon a chosen set of information sources, some maintained locally and some accessed remotely through the World Wide Web. A particular emphasis of START is that of providing *high-precision* information access, such that the user may maintain a fair degree of confidence that a response, if returned by the system, is appropriate to the submitted question. This is particularly useful in the context of mobile devices with small screens, whose users would much prefer seeing short, high-precision responses with "just the right information" rather than long lists of possibly-relevant materials.

When an English question is received by START, the question is parsed and translated into a set of *nested ternary expressions*, which are constituent–relation–constituent triples that capture much of the most salient information in a syntactic parse tree in a form well-suited to matching [1]. Other information contained within the question is stored as ancillary information and is used only secondarily in matching, providing efficiency without information loss.

In order to obtain responses to user questions, START matches analyzed questions to *natural language annotations*—descriptive English phrases and sentences that have been associated with retrievable segments of information or with subsets of parallel information within information sources [2]. START's natural language annotations are also processed into nested ternary expressions, and the matching occurs on this level, assisted by an array of techniques that include matching through synonymy and hyponymy, the application of structural transformation rules [3], a reference resolution mechanism, a mechanism for matching terms to

parameters that represent source-specific semantic categories [4], and a mechanism for decomposing complex questions syntactically into subquestions for independent processing [5]. Once a question has been matched to a natural language annotation, the information segment or information source associated with that annotation can be accessed to provide a response to the user.

Figures 2 and 3 present screenshots from the publicly available START server; such questions can also be asked from mobile devices.

Figure 2. START handling a comparative question about countries' areas

Figure 3. START performing a currency conversion

Further information on the operation of the START system can be found in [2] and [4], and in [6] (also presented at ICSC 2007).

3. From information access to actions

START is used for general information access in the StartMobile system, and START's answers to such questions are simply relayed to the user's mobile device for presentation to the user. Figure 4 illustrates this capability used to retrieve multi-media information.
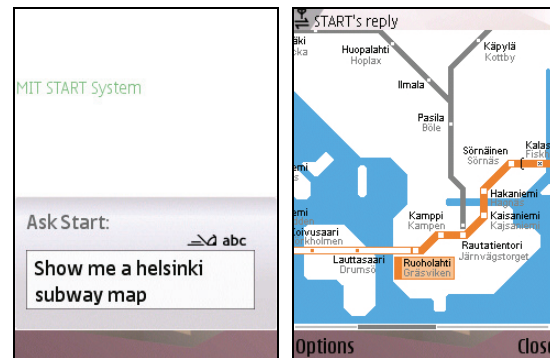

Figure 4. StartMobile responding to a request for general information

Figure 4 depicts a fully grammatical request submitted by the user. As part of the StartMobile effort, we have extended START to allow the user to enter not only grammatical variants of requests, but also fragmentary utterances in a range of cases where the meaning remains clear. For the request illustrated in Figure 4, for example, the user could have entered "Helsinki subway map" or "subway map Helsinki", resulting in a display of the same map.

Other types of requests concern information maintained on the user's mobile device. To handle these requests, START matches them to natural language annotations as always; however, the annotated material in this instance is a procedure that relays instructions to the mobile device. Associated software on the mobile device performs the necessary operations and delivers the results to the user. Figure 5 depicts the handling of such a request, involving a search through the calendar on the user's mobile device.
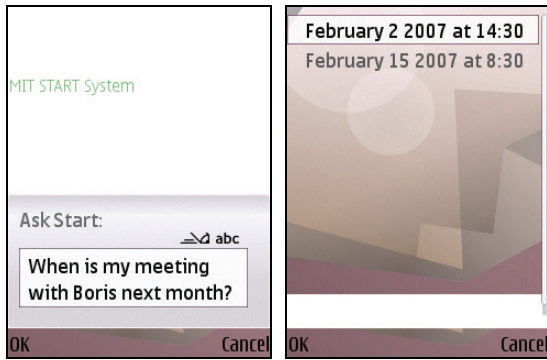
Figure 5. StartMobile responding to a request by searching through the calendar on the user's mobile device

Requests that concern information on a user's mobile device may contain unusual names—people, streets, cities, businesses, etc.—that appear within particular data entries on the user's mobile device. To enable START to correctly analyze these requests and take appropriate actions, StartMobile implements a special mechanism whereby submitted user requests are initially inspected, on the user's mobile device, to recognize and categorize names that appear in data sets such as the contacts database or calendar. This information is then utilized in subsequent analysis and processing of those requests.

In another set of cases, the user's input is not a request for information, but rather a command to perform an action on the user's mobile device. These requests are handled in a similar manner to requests for information on the user's mobile device, with START relaying instructions to software that performs actions on the user's mobile device. Figure 6 illustrates StartMobile's handling of a request to take a picture using the camera on the user's mobile device.
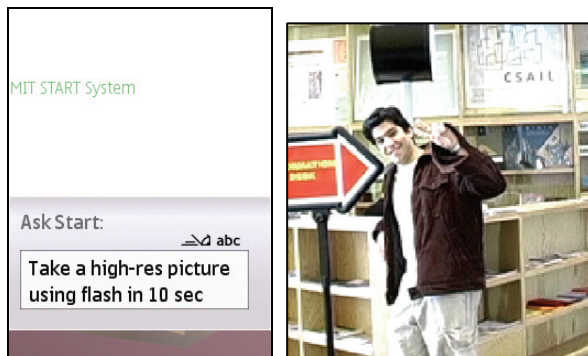

Figure 6. StartMobile responding to a request to take a picture using the phone's camera

In still other cases, the user may enter a request on one mobile device to perform an action on another mobile device. In this instance, START will relay instructions to the first device, which must then relay appropriate instructions to the affiliated device. Figure 7 presents an example of this type of request being handled by StartMobile.
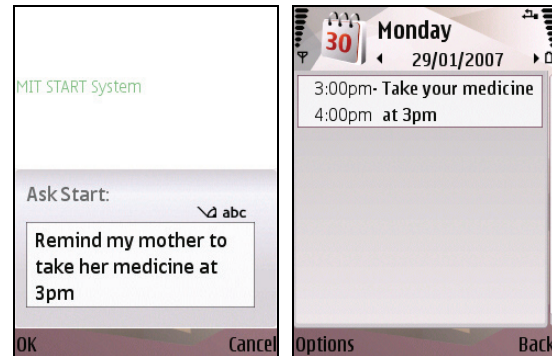

Figure 7. StartMobile responding to a request to post a reminder on an affiliated mobile device

As stated above, a key goal of the StartMobile system is to accept a full range of natural language variants for each request type handled by the system. This is accomplished in large part through the use of natural language annotations. START's parameterized annotations and structural transformation rules make it possible for a single annotation to cover many requests, reducing the burden of writing annotations to a manageable task. START's natural language interface frees the user from having to memorize particular ways to phrase requests. Furthermore, this association of user requests with system capabilities is accomplished with very high precision: it is extremely rare that a user question will be matched incorrectly to a request type handled by the system.

4. Handling ambiguous requests

When natural language requests are to be processed in dynamic, distributed environments such as those involving mobile devices, an additional problem arises. It is often the case that natural language requests can only be fully understood—their ambiguities resolved—in the presence of specific, matching components of knowledge, and in distributed environments, this knowledge is distributed, requiring the networked devices and systems in some cases to collaborate not only toward the ultimate satisfaction of the received requests, but as well toward the initial *understanding*

of those requests, so that it may be possible to satisfy them.

In the course of the StartMobile effort, we have developed an intermediate representation language, called Moebius, that supports distributed interpretation and distributed fulfillment of natural language requests. Moebius serves to encode natural language requests at varied stages of syntactic and semantic interpretation, so that these requests may be relayed between systems—for instance, a user's mobile device, central servers, and other users' mobile devices—in order to receive additional interpretation and fulfillment. While Moebius specifically addresses the representation and processing of ambiguous requests, it is also applicable for more straightforward requests and thus we use the language as an intermediate representation for all StartMobile requests that must be relayed between systems and devices.

Following is an example of a Moebius expression issued by START to the user's mobile device, depicting a substantially interpreted version of the English request illustrated in Figure 7.

```
alert(object:person mother(of:person "user"),
     with:message_string
        "Take your medicine at 3pm.",
     at:time "2007-01-29T15:00:00")!
```

Moebius specifies syntactic relations loosely along the lines of a *dependency tree* representation (see, for example, [7]), and it adds semantic labels, drawn from a hierarchy of general to specific categories, to various elements within the representation.

A key aspect of Moebius is that it uses language itself as a representation. In this respect, it shares a common orientation with the START system. START uses language-based ternary expressions as a representation for both questions and natural language annotations. Indeed, when START matches a question to a natural language annotation, it does two things: it provides an answer to the question, and it commits to an interpretation of the question. Moebius can be thought of as extending this idea to distributed contexts, enabling partially-interpreted requests to be interpreted and fulfilled by collective action on the part of multiple systems in a distributed environment.

Also in using language as a representation, Moebius shares a common orientation with work in computer-processable languages, such as the work on ACE [8] and CPL [9]. Whereas these approaches use simple language as a medium for translating knowledge into more formal representations, however, Moebius uses simple language for its ability to express and preserve the same sorts of ambiguities that arise in full-English requests.

As an example of the use of Moebius to characterize an ambiguous request at different stages of interpretation, consider the request

    Is Carl at IBM?

This question could be offered to ascertain whether or not Carl is employed by IBM, or it could be offered to determine whether or not Carl is, at the moment, physically present at an IBM facility. Which interpretation is intended may be construed by consulting the repertoire of capabilities offered by the device expected to fulfill the request—whether that device is known to be able to respond to one interpretation or the other—or, the interpretation may be construed by reference to contextual information from the current state of processing, or by consultation with a human user.

If the device that initially processes the request "Is Carl at IBM?" does not have access to the knowledge needed to fully interpret the request, then, using Moebius, that device can encode the request in a partially interpreted form:

    be(subject:person "Carl", at:object "IBM")?

This representation parses the request syntactically, yet it makes no commitment as to the semantic interpretation of the relationship between "Carl" and "IBM" or as to the specific semantic category of "IBM" ("object" being the most general semantic category). If this request is relayed to another device or system that possesses the necessary knowledge to disambiguate the request, that system may cast the request into one of two more fully interpreted forms. If the determination is made that the request concerns physical presence at an IBM facility, the request can be re-expressed as

    be(subject:person "Carl", at:facility "IBM")?

where "IBM" is classified semantically as a physical "facility". On the other hand, if the determination is made that the request concerns employment, the request can be re-expressed as

```
employ(subject:organization "IBM",
       object:person "Carl")?
```

where "IBM" is classified semantically as an abstract "organization" and the relationship is re-expressed as one of employment. Subsequent processing can continue, then, according to the chosen interpretation.

The present version of StartMobile does not have the ability to determine current locations of individuals, yet it can access employment information stored in the contacts database of the user's mobile device. For this reason, the mobile device, upon receiving the partially interpreted form of "Is Carl at IBM?" from START, simply translates this expression into the more fully-interpreted Moebius expression requesting employment information for Carl, then processes this Moebius expression. Figure 8 illustrates StartMobile's handling of the request "Is Carl at IBM?", given appropriate employment information recorded in the contacts database of the user's mobile device. The displayed output in this case explicitly informs the user that StartMobile has retrieved employment information, so as to clarify StartMobile's interpretation of the user's request.
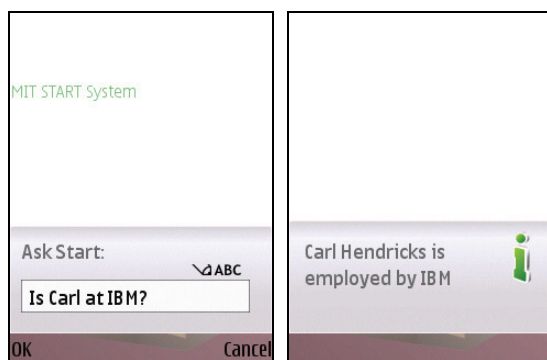

Figure 8. StartMobile responding to an ambiguous request

On first glance, it might appear that StartMobile's interpretation of the request "Is Carl at IBM?" is less than well motivated. However, on closer examination, this interpretation is indeed well motivated if considered in the context of assumed conversational maxims such as proposed by Grice [10]. These maxims exhort the individual composing an utterance in a conversational context to abide by conventions in supplying an adequate amount of information, but not too much information, by being truthful, by supplying only relevant information, and by being clear or perspicuous. If it may be assumed that the user of StartMobile is generally aware of the system's capabilities, such as being able to obtain employment information but not being able to determine current locations of individuals, then it is reasonable to assume that the user, when generating a request such as "Is Carl at IBM?", does so with the full expectation that the

request will be fulfilled by accessing employment information. From the system's standpoint, then, it is safe to proceed by assigning an interpretation dictated by that system's capabilities, on the assumption that the user is aware of those capabilities and has constructed his or her request in compliance with accepted conversational maxims.

In general, ambiguity can arise from many sources—abstract verbs; syntactic ambiguities; omitted adverbial phrases; ambiguous prepositions and conjunctions; abstract semantic categories; descriptive specifications of objects; ambiguous names, times and places; anaphora; and abstract adjectives and adverbs, to name a few. Moebius provides several mechanisms for depicting and resolving these ambiguities. Abstract verbs can be replaced by more specific verbs. For example, a request to "contact" a person can be re-expressed as a request to "call" a number or "send" a message. Abstract semantic categories can be replaced by more specific categories. For example, "message" can be replaced by "e-mail_message", "text_message", "voice_message", etc. Descriptive sub-expressions such as "address(of:facility apartment(of:person "Sandra"))" and "3 o'clock" can be replaced by more specific expressions such as "298 Beacon Street, Boston, MA 02116" and "2007-04-22T15:00:00". In addition, ambiguous commands, statements and questions can be clarified by inserting adverbial phrases, for example, or replacing the original expressions with entirely different expressions. The goal of Moebius is to capture a range of such ambiguities in various stages of interpretation, and we have found that simple language, structured for ease of computer processing, provides sufficient expressiveness to model many common ambiguities.

5. Alternate configurations and capabilities

The present StartMobile system situates START as a central server, accessed by one or more mobile devices which occasionally interact directly with one another. START performs initial processing of user requests, then passes Moebius requests to other systems and devices as needed for further interpretation and/or fulfillment. However, the overall design of the StartMobile system allows for other configurations as well.

One alternative is to perform initial processing of natural language requests on each user's mobile device. This alternative would appear to be quite feasible given recent enhancements to processor speeds and available memory within mobile devices. Each mobile device could then relay Moebius requests

to other devices and systems as necessary for further interpretation based on knowledge held by those systems, or for the completion of actions external to that mobile device.

A third possibility involves a mixture of these two approaches. In this configuration, each user's mobile device would contain a lightweight capability for simple language processing, then pass off partially-interpreted requests or even uninterpreted requests to other, more substantial language processing components as the need arises.

Additional request-handling capabilities are also supported by the overall design of the StartMobile system. Currently with StartMobile, mobile devices and central servers are networked to access information on either type of system or to perform actions on the mobile devices. However, other types of components can be included in the mix: peripherals such as keyboards and printers, proximity-sensitive servers (e.g., servers associated with museums or businesses), software agents and other components, and additional types of actions can be introduced that involve any of these systems (setting or reading a home thermostat, for example). Because these added participants and capabilities arrive, in all cases, with vocabulary and turns of phrase to describe them in language, it is our expectation that language-based matching of requests to annotations and the language-based Moebius encoding can be expanded accordingly to accommodate these entities.

6. Continuing research

We are continuing to add new capabilities to the StartMobile system. In particular, we are focusing our current efforts on the handling of requests that involve aggregation of general information, social networking, context awareness (e.g., location and schedule awareness), finer control of mobile device features, and interaction with peripherals and other devices.

In parallel, as we incorporate new capabilities into StartMobile, we expect to encounter additional varieties of ambiguous requests. While the Moebius language has been kept simple as a matter of design, our intention is to expand its features incrementally as required to represent these new types of ambiguity.

A separate path of enhancement lies in the incorporation of additional modalities of language-based input. Speech input is certainly a strong candidate for coupling with the techniques presented in this paper and could be particularly useful in cases that do not involve noisy environments, a need for privacy, or excessively broad vocabularies. Touch screen interactions could also be used to streamline the input of language-based requests to mobile devices. While the use of language, in particular, as a vehicle for specifying wide-ranging requests to mobile devices is certainly a powerful notion, the exact means by which language-based requests are entered to a device need not be restricted to any single modality.

7. Contributions

Cell phones and other mobile devices have the potential to provide users with many useful features and capabilities, but the more capable these devices become, the harder it becomes to make use of them with only buttons and menus. Our StartMobile prototype demonstrates how natural language, as an alternative access method, can provide convenient access to complex and even abstruse features.

There are two principal contributions of this work. The first is a demonstration of utility on the part of language to increase the usability of mobile devices. This was accomplished by adapting the START information system for use with mobile devices, developing the Moebius language for the representation of partially-interpreted requests, and creating a body of software that fulfills a range of requests on mobile devices.

The second contribution of this work is the overall design of the StartMobile system. This design provides flexibility with respect to configurations and capabilities supported, due to its reliance on language-based representations—in particular, natural language annotations and the Moebius representation language—and due to its explicit representation and handling of partially-interpreted natural language requests in the distributed context of mobile environments.

8. Acknowledgements

9. References

[1] B. Katz, "Using English for Indexing and Retrieving," in *Artificial Intelligence at MIT: Expanding Frontiers, v. 1*, Cambridge, Massachusetts, 1990.

[2] B. Katz, "Annotating the World Wide Web Using Natural Language," *Proceedings of the 5th RIAO Conference on Computer Assisted Information Searching on the Internet (RIAO '97)*, Montreal, Canada, 1997.

[3] B. Katz and B. Levin, "Exploiting Lexical Regularities in Designing Natural Language Systems," *Proceedings of the 12th International Conference on Computational Linguistics (COLING '88)*, Budapest, Hungary, 1988.

[4] B. Katz, G. Borchardt, and S. Felshin, "Natural Language Annotations for Question Answering," *Proceedings of the 19th International FLAIRS Conference (FLAIRS 2006)*, Melbourne Beach, Florida, 2006.

[5] B. Katz, G. Borchardt, and S. Felshin. "Syntactic and Semantic Decomposition Strategies for Question Answering from Multiple Resources," *Proceedings of the AAAI 2005 Workshop on Inference for Textual Question Answering*, Pittsburgh, Pennsylvania, 2005.

[6] B. Katz, G. Borchardt, S. Felshin, Y. Shen, and G. Zaccak, "Answering English Questions using Foreign-Language, Semi-Structured Sources," *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, Irvine, California, 2007.

[7] J. Carroll, T. Briscoe, and A. Sanfilippo, "Parser Evaluation: a Survey and a New Proposal," *Proceedings of the International Conference on Language Resources and Evaluation*, Granada, Spain, 1998.

[8] N. E. Fuchs, K. Kaljurand, and G. Schneider, "Attempto Controlled English Meets the Challenges of Knowledge Representation, Reasoning, Interoperability and User Interfaces," *Proceedings of the 19th International FLAIRS Conference (FLAIRS 2006)*, Melbourne Beach, Florida, 2006.

[9] P. Clark, P. Harrison, T. Jenkins, J. Thompson, and R. Wojcik, "Acquiring and Using World Knowledge using a Restricted Subset of English," *Proceedings of the 18th International FLAIRS Conference (FLAIRS 2005)*, Clearwater Beach, Florida, 2005.

[10] H. P. Grice, "Logic and Conversation," *Syntax and Semantics, Volume 3: Speech Acts*, P. Cole and J. L. Morgan editors, Academic Press, 1975.