Context-Based Information Access

Jamie Callan
Carnegie Mellon University
Pittsburgh, PA 15213
callan@cs.cmu.edu

Nick Belkin
Rutgers University
New Brunswick, NJ 08901
nick@belkin.rutgers.edu

Executive Summary

The continued growth of traditional forms of information, the recent rapid growth of sensor information, and the growing use of information in mobile and other novel applications leads researchers in a variety of Computer Science research areas to identify context-based information access as a research priority in the next decade. "Context" covers a broad range of topics ranging from how and where data was collected to why the information is wanted and how and where it will be used. Although researchers in information retrieval, databases, and sensors view the problem from very different perspectives, they agree that contextual information is a key to improved use and reuse of information in the coming decade. Research topics include context abstractions; detecting, sensing, and discovering context; using context; manipulating context information; and interpreting context. This research area is likely to require diverse teams doing integrative research, with consequently larger funding needs. Important obstacles to research in this area are the current lack of tools for collecting contextual information, and the current lack of generally available datasets and testbeds for conducting and comparing research.

Introduction

Most information access tools are designed for ad-hoc, single-use interactions with anonymous individuals. However, people now use computers routinely, for a wide variety of tasks, in a wide variety of settings, over multi-year periods. People also assemble ad-hoc personal digital libraries that they carry with them from one computer to another, across a lifetime of use. There is growing recognition that the next generation of information access tools must make greater use of context, in the form of detailed user, task, and situational models, to provide greater personalization and improved accuracy [Allan and Croft, 2003; Cool and Spink, 2002; Finkelstein, et al., 2002; CAMM 2002; IMUIP 2003].

Scenarios

The role of understanding context to improve information access is evident in a wide variety of settings. The brief scenarios below are meant to indicate just some of these.

- Collaboratories. In collaboratories, the partners in the enterprise must communicate with one another in both face-to-face and distance modes, and must be able to integrate results and ideas from a wide variety of participants over extended periods of time and across multiple locations. In order for them to communicate and work effectively, and in particular to access and use appropriate information, their information and communication support systems must have knowledge of such aspects of all of their contexts as:

  o Past work,
  o Current work,
  o Location,
  o Role in the collaboratory

- Monitoring. Ubiquitous computing, with sensors distributed throughout the environment, will lead to a new class of situations in which knowledge of the context will be of primary importance. The data from sensors, to be properly interpreted and used, will need to be integrated with specific information about the context in which they exist, whether it is a personal context, such as an individual's medical history, or a physical and functional context, such as a process control environment. It is also important in all such situations, that the context of the sensor itself (e.g. local temperature, humidity, past behavior) be understood and used in interpreting its data.

  o Personal monitoring
    ▪ Health, family, patients
  o Environments
    ▪ Power plants, hazardous environments, remote environments
  o People in mobile situations.

  Knowledge of a person's location and movements will be easily determined through networks of sensors and GPS devices. Such contextual information will be useful in determining the most appropriate forms of information system support at any particular point in space-time. Yet this form of personalization cannot be accomplished without deep knowledge and understanding of other aspects of the person's context, such as preferences for particular types of interaction, salience and urgency of a message, and types of interaction/communication devices available.

This report is intended to reflect the consensus of a group of researchers who met at the 2003 NSF Information and Data Management (IDM) Principal Investigator (PI) workshop to discuss greater use of context for information access. Although the researchers shared a common interest in context, they came from three relatively distinct research communities (Information Retrieval, Databases, and Sensors). At an abstract level these communities tend to agree about what is important and what the research priorities ought to be. It is difficult for these communities to agree at a more specific or detailed level due to their differing research agendas.

Context in Information Access

There is broad agreement that context is important and not sufficiently exploited by current information systems. Contextual information can be gathered about and applied to communities (e.g., demographics, social contexts), individuals (e.g., tasks, goals, previous and ongoing behaviors, location), locations (e.g., characteristics of the physical environment), data, and infrastructure. Context means different things to different research disciplines, but even within a specific discipline it would be helpful to develop taxonomies of context types.

Using contextual information for information access involves addressing five issues, described below.

- **Context abstractions:** In order for researchers to develop general theories it is necessary to abstract from the details of a particular situation to a set of characteristics that are more broadly applicable. One research priority is to identify useful abstractions of contextual information that permit such generalization. Candidates that are often mentioned include static vs. dynamic, local vs. global, and short-term vs. long-term. A related problem is the study of how to *represent* and *store* contextual information for reuse.

- **Detecting, sensing, and discovering context:** Researchers distinguish between *explicitly-derived context*, which may relatively easy to acquire, and *implicitly-derived context*, which may be more difficult to acquire. *Sensors* of many types will be essential to acquiring explicit contextual information. Implicit contextual information is acquired via passive observation, for example of a person's behavior, and improved *inference or learning* will be required to acquire it. *Privacy* is a serious concern, from both technological (e.g., what it is possible to protect) and policy (e.g., what it is desirable or legal to protect) perspectives.

- **Using context:** Once contextual information is acquired, how can or should it be used? The priority for this group was to use it for *improving information access*, such as improved ease of use, better user interfaces, improved accuracy, information that is more suitable for the current information need, personalized information, and information tailored for a particular use. *Privacy* protection is required to control what information can be used in a given situation. *Context-sensitive resource optimization* enables devices and software to respond more effectively to current priorities and environmental characteristics, as well as to assist in configuration and to monitor maintenance needs. Contextual information can support more fine-grained methods of *access control*. Finally, when a wealth of contextual information is available, *selecting the appropriate contextual information* to save and use in a given situation becomes an important problem.

- **Manipulating context information:** Making contextual information a more central part of information access and delivery will make it increasingly valuable, encouraging research on reuse and repurposing of context information. *Context management*, and methods of *mapping, comparing, transforming,* and *combining* contexts will be required.

- **Interpreting context:** A central issue for contextual information access is how the contextual information is to be *interpreted* by the support systems in order to improve information access. Interpretation of physical data concerning the context of sensors in order to assign credibility is a relatively simple example of this general problem. Inferring a person's current goal from their current and past behaviors, and from the details of their current activity, and using the results of the inference in order to personalize the interaction to suit that goal, is substantially more complex.

There are important research issues associated with each of the above aspects of context. The specific research issues depend upon the type of context and situations, but the general requirements span the range of context types and specific information access applications.

Research Directions

The research community studying the use of context for improved information access is likely to be more diverse than other research communities. There are many research problems that can be studied by a single investigator working within a single research community, but many of the interesting problems will require teams with varying expertise, for example, teams that include sensor experts, application experts, and researchers from the cognitive and behavioral sciences. Such diverse teams are difficult to assemble, are even more difficult to coordinate, and have substantially larger funding requirements (e.g., in the $800K - $1.5M range).

Two particular needs for this line of research are *tools for collecting contextual data* and *widely available testbeds and datasets* of contextual data (e.g., TREC). The importance of widely available tools, testbeds and datasets to this community cannot be overstated; without them, each research team wastes a significant fraction of its budget reinventing the wheel, and lack of such resources is a serious barrier to entry for young researchers and researchers from less-established universities. *Libraries of open-source software* for important algorithms would also be helpful. A larger number of funded projects in this research area would help build critical mass and might spur development of community resources.

As in any new research area, there is a need for significant fundamental research. Particularly important is research that provides framework for integration across user-focused and device-focused contexts, and integration across domains.

Recommendations

NSF must make a substantial long-term commitment to establish, and particularly to *maintain*, datasets, testbeds, and open-source software libraries to support research on context-based information access. The cost of acquiring contextual information is now simply too high and requires too much expertise for most research groups to acquire the kind of information that will be available routinely in the next decade. Requiring each research group to acquire such information on its own poses a high barrier to entry and also makes research results nearly impossible to compare across research groups. NSF must face this problem and make a long-term commitment to solving it, possibly in conjunction with other government organizations, such as NIST, or professional societies such as ACM or IEEE. An initial step in this direction might be to initiate and sponsor a series of workshops, perhaps in partnership with professional societies.

NSF should also take on a greater role in fostering cooperation between other governmental agencies and the research community. Many government agencies collect large amounts of sensor or other contextual information that would be useful for research purposes. NIST, NIH, DoD, the intelligence agencies, and NIMA are just a few examples of potential partners to the research community. Repurposing government data for research use requires directly confronting the privacy and context manipulation research issues mentioned above, but these must be faced eventually anyway. The government operates some of the biggest digital libraries in the world, and tens of thousands of people use those digital libraries every day. These are a vast, untapped resource, comparable to what the commercial sector uses. Providing access to researchers, in some form, should be an NSF priority.

References

- Allan, J. and Croft, B., editors. (2003) Challenges in information retrieval and language modeling. SIGIR Forum, v. 37, no. 1: 31-48.
- Cool, C. & Spink, A. (2002) Special Issue on Issues of Context in Information Retrieval. *Information Processing & Management*, v. 38, no. 5
- Finkelstein, L. et al. (2002) Placing search in context: the concept revisited. *ACM TOIS*, v. 20, no. 1: 119-131.
- *NSF Workshop on Context Aware Mobile Database Management (CAMM)*. January 24-25, 2002. http://www.sice.umkc.edu/nsfmobile/wshop.html/
- *SIGIR 2003 Workshop on Implicit Measures of User Interests and Preferences*. August 1, 2003. http://research.microsoft.com/~sdumais/SIGIR2003/SIGIR2003-ImplicitWorkshop.htm