

SDP 10 Sensorbase: A Story in Three Parts

SDP 10.1 People

- Principal Investigator: Mark Hansen
- Faculty: Mark Hansen (UCLA Statistics, CENS Area Lead - Statistics), Deborah Estrin (UCLA Computer Science, CENS Director)
- Staff: Richard Guy (UCLA Computer Science, CENS Staff)
- Graduate Students: Keith Mayoral (UCLA Computer Science, CENS GSR)

SDP 10.2 Overview

Sensorbase provides a database backbone infrastructure for many of the sensing campaigns and projects that take part at CENS. It provides a mean to store all sensor data gathered in the field in a user friendly environment through the web interface, as well as providing an automated upload process using SOAP and RESTful services. With it, users can store any type of sensor data, be it text, audio, video or anything which can be represented in binary. To set itself apart from a simple relational database, Sensorbase also provides a user friendly web-based method of viewing and editing stored data, as well as a concept of a user's data. Keeping all data in a centralized location provides the benefit of allowing users to share pertinent datasets with other researchers. To ensure that only people who should see your data can, users can also specify the specific types of permissions to apply to each dataset or subset of. By providing a standard programmatic approach to retrieving data, users can process and manipulate their data straight from the Python or PHP scripts that use the data.

SDP 10.3 System Description

Sensorbase, as mentioned before, is designed to be an online accessible system in which researchers can upload or *slog* sensor data to our central server running a MySQL instance.

The main interface to this system is through the website Sensorbase.org. The structure of the web interface is as follows:

Users must log in to the system through the front page of the site. No data or information related to any projects can be viewed by an anonymous visitor to the site. Each applying prospective user must submit an email declaring his or her area of research and how Sensorbase will be of use. Once reviewed, a user account is manually created for each applicant. This process ensures that only researchers have access to the data available through the website. Once a user has registered, she can start creating datasets for each project related to her research.

A **project** in Sensorbase is a collection of tables all related to a single real world research project. Within each project, individual **tables** are defined to further separate different datasets within a research project. These tables can contain any number of fields of varying types, from simple text and numbers, to images and sounds. This structure forms the basis of how datasets are stored and viewed in Sensorbase.

Data Permissions and privacy concerns are also managed through the website interface. There are four different types of permissions which can be granted either at the project level or at the finer table level. If a user has a greater permission granted to the project containing a table, then the project level permission supersedes the table level permission.

Build access gives a user read/write/create/destroy privileges on the specified project or table. Project creators start with build access, but can also grant it to other users or *collaborators*.

Slog and read access gives a user the read/write ability on the specified project or table. This is the equivalent of assigning someone as a *contributor* to your project without actually having control over the project configuration itself.

Slog only access is given to a user if you don't want them to have read privileges, but still require them to contribute data. This setting could apply in situations where you have users submitting personal information, like GPS location traces throughout the day, which would be considered private and should only be seen by the researcher(s) in charge of the project.

Finally, **Read only access** can be granted to individual users who you might want to share collected data with, but have no association with the project/table such that they don't need access to write or modify the data.

The permission levels described above define how specific portions of the data can be shared. Another privacy setting which applies to a whole project lets you declare whether the data collected is public or private. This affects whether or not an arbitrary user can search for and browse your data from the search functionality present in the website interface. If a user attempts to view your project when marked as private, they will be redirected to a page stating that access is not allowed. Projects marked as public automatically grant read access to all users of the website.

SDP 10.4 Accomplishments

Sensorbase has had many improvements over the few months which have helped it become even more useful as an online database.

Integration of code into the main Sensorbase build was done at the start of the year to incorporate different functionality which was developed by other CENS researchers in individual projects. John Hicks, a researcher at CENS added an Abode Flash based data graphing feature and a delayed query feature to allow for large dataset dumps to be scheduled to run at times when the site is not under heavy load. These features, as well as smaller cosmetic changes, were added to Sensorbase prior to the addition of SVN access described below.

Programmatic control of datasets has been one type of feature which was requested by Sensorbase users. While support for the programmatic retrieval of data already existed in Sensorbase, it became apparent that other features would be as helpful if there were implemented for programmatic access. These features, which were implemented as RESTful services, allow users to programmatically deal with projects and/or tables given that they already have proper permissions. Users can now create, update, delete, or modify the structure of a specified project or table. Furthermore, more advanced SQL queries are available through this programmatic interface, such as table joins. For a complete list of implemented features, one can visit the help page at <http://sensorbase.org/help/> as well as see example usage of each function.

Availability of source code is another aspect of Sensorbase which we've worked to simplify. We believe that the best way to have Sensorbase work for all users is to allow them to modify the base code to better suit their project's individual needs. If a feature which has been implemented in one of these separate projects can be generalized to work for other groups, then we can simply transfer the code over to the main distribution to make it available to all users.

In the past, when a group requested a copy of the Sensorbase source code, they would simply get an archive containing a prepackaged version of Sensorbase to set up on their own. We have recently set up an anonymous SVN repository which allows for anonymous read access of the source code for the current stable version of Sensorbase. When users download the source, they can find a text document to assist in the setup of their own Sensorbase build. By providing this resource, we can allow outside developers to configure Sensorbase for their own uses as well as contribute to the improvement of our main build.

Improving database speed and storage has been a large milestone for this year. We started the process of transferring over the underlying MySQL database and the vast directories containing all the binary data associated with Sensorbase to a much larger Sun Sunfire server named Thumper. This server currently has 24TB of space so sufficient space for data storage won't become an issue anytime soon. Nonetheless, we have gone about restructuring the directory structure of the folders which contain all the binary data such as images and sounds. This was due to the fact that on the old server, there was considerable delay added to file access because of the

number of files stored in each directory. By restructuring the database and separating the database backend from the UI frontend, we hope to increase the overall speed of queries and improve the performance of the site overall.

Many other smaller bugs and fixes have happened over the year as they have been pointed out by the users of Sensorbase which as a whole keep us just as busy as the larger issues. Users can always send requests for features or bug fixes to sb-support@lecs.cs.ucla.edu.

SDP 10.5 Future Directions

Much of the work planned for the coming year is based on what features the research groups at CENS will be requiring for their respective projects. We have been in talks with groups from all parts of CENS to assess the requirements each group will need for their projects to interact with Sensorbase effectively. Some of the groups involved in talks include: CENS Deployment Center, Urban Sensing, NetworkedNaturalist, PeruNet, as well as others. The features they request will have highest priority in terms of new features to implement.

Adding a layer on top of Sensorbase which interacts with the database to notify users of configurable alerts is also a feature we will be implementing in the coming year. Allowing for an email to be sent to a user when a data entry passes a threshold or a project attribute changes would help users be better informed about the data entering their projects, especially when much of the data is automatically being slogged by the sensing devices themselves.

Summary

As stated in the sections above, many research groups both in CENS and out, use Sensorbase to store the data collected in their studies. Sensorbase plays a vital role in simplifying the process of organized data collection, processing, and dissemination. By continuously expanding upon the feature set and usability of the system, we can provide ongoing support for the sensing research community.

