SIDGrid: A Framework for Distributed, Integrated Multimodal Annotation, Archiving, and Analysis

Gina-Anne Levow, Sonjia Waxmonsky
Department of Computer Science
University of Chicago
levow,wax@cs.uchicago.edu

Bennett Bertenthal
Department of Psychology,
Indiana University, and
Computation Institute
University of Chicago
bbertent@indiana.edu

David McNeill
Department of Psychology
University of Chicago
dmcneill@uchicago.edu

Mark Hereld, Sarah Kenny, Michael E. Papka
Computation Institute
The University of Chicago
m-papka,m-hereld,skenny@uchicago.edu

Abstract

The SIDGrid architecture provides a framework for distributed annotation, archiving, and analysis of the rapidly growing volume of multimodal data. The framework integrates three main components: an annotation and analysis client, a web-accessible data repository, and a portal to the distributed processing capability of the TeraGrid. The architecture provides both a novel integration of annotation, analysis, and search for multimodal data and a powerful framework for web-based, distributed collaborative annotation and analysis. The flexibility and capabilities of the system have been demonstrated through archiving Talkbank and other spoken discourse and dialogue data and performing joint multimodal analysis of lexical, prosodic, turn-taking, and other multimodal factors.

1 Introduction

Recent research programs in multimodal environments, including understanding and analysis of multi-party meeting data and oral history recording projects, have created an explosion of multimodal data sets, including video and audio recordings, transcripts and other annotations, and increased interest in annotation and analysis of such data. However, multimodal data poses particular challenges including a broad range of annotation and analysis measures, large storage requirements for media data, and increased computational complexity of media data and multi-factor analyses. Furthermore, since this data is costly to collect and annotate, both in terms of time and money, there is additional incentive to share data and collaborate on annotation efforts. The wide range of annotations, from aligned transcripts to gaze to reference to gestural form, often leads to annotation by multiple expert groups, possibly geographically distributed, to fully exploit these resources.

A number of systems have been developed to manage and support annotation of multimodal data, including Annotation Graphs (Bird and Liberman, 2001), Exmeralda (Schmidt, 2004), NITE XML Toolkit (Carletta et al., 2003), Multitool (Allwood et al., 2001), Anvil (Kipp, 2001), and Elan (Wittenburg et al., 2006).

The framework described here, developed under the NSF Cyberinfrastructure Program, aims to extend the capabilities of such systems by focusing on support for large-scale, extensible distributed data annotation, sharing, and analysis. The system is open-source and multi-platform and based on existing open-source software and standards. The system greatly eases the integration of annotation with analysis though user-defined functions both on the client-side for data exploration and on the TeraGrid for large-scale distributed data processing. A web-accessible repository supports data search, sharing, and distributed annotation. While the framework is general, analysis of spoken and multi-modal discourse and dialogue data is a primary application.

The details of the system are presented below. Sections 2, 3, and 4 describe the annotation client,

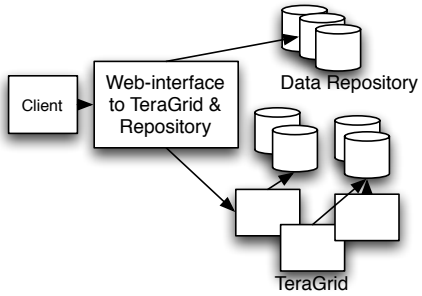

Figure 1: System Architecture

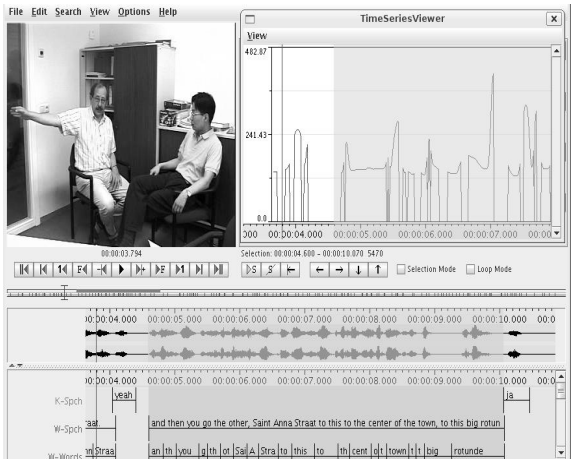

Figure 2: Screenshot of the annotation client interface, with video, time-aligned textual annotations, and time series displays.

the web-accessible data repository, and the portal to the TeraGrid, respectively, as shown in Figure 1 below. Section 6 describes system availability and planned extensions to system functionality.

2 The SIDGrid Client

The SIDGrid client provides the primary interactive multimodal annotation interface. A screenshot appears in Figure 2. The client extends the open-source ELAN annotation tool from the Max Planck Institute[1]. ELAN supports display and synchronized playback of multiple video files, audio files, and arbitrarily many annotation "tiers" in its "music-score"-style graphical interface. The annotations are assumed to be time-aligned intervals with, typically, text content; the system leverages Unicode to provide multilingual support. Time series such as pitch tracks or motion capture data can be displayed synchronously. The user may interactively add, edit, and do simple search in annotations. For example, in multi-modal multi-party spoken data, annotation tiers corresponding to aligned text transcriptions, head nods, pause, gesture, and reference can be created.

The client expands on this functionality in two main ways. First, the system allows the application of user-defined analysis programs to media, time series, and annotations associated with the current project, such as a conversation, to yield time series files or annotation tiers displayed in the client interface. Any program with a command-line or scriptable interface installed on the user's system may be added to a pull-down list for invocation. For example, to support a prosodic analysis of spoken dialogue data, the user can select a Praat (Boersma, 2001) script to perform pitch or intensity tracking. Currently a variety of Praat, R, and Matlab scripts are supported, and topic segmentation and reference resolution algorithms are being integrated. Also, the client provides integrated import and export capabilities for the central repository. New and updated experiments and annotations may be uploaded directly to the archive from within the client interface. Existing experiments may be loaded from local disk or downloaded from the repository for additional annotation.

3 The SIDGrid Repository

The SIDGrid repository provides a web-accessible, central archive of multimodal data, annotations, and analyses. This archive facilitates distributed annotation efforts by multiple researchers working on a common data set by allowing shared storage and access to annotations, while keeping a history of updates to the shared data, annotations, and analysis.

The browser-based interface to the archive allows the user to browse or search the on-line data collection by media type, tags, project identifier, and group or owner. A simple permission scheme, based on Unix-style group permissions, provides public access to freely available data while restricting access to more sensitive data to authorized users. Once se-

[1] http://www.mpi.nl/tools/elan.html

lected, all or part of any experiment may be downloaded. In addition to lists of experiment names or thumbnail images, the web interface also provides a streaming preview of the selected media and annotations, allowing verification prior to download. (Figure 3)

The repository also supports import of new data. To support interoperability with other annotation tools, conversion functions have been developed for a range of annotation formats, in collaboration with developers[2], using Annotation Graphs as an interchange format, in addition to the existing ELAN-based import capabilities.

All data is stored in a MySQL database. Annotation tiers are converted to an internal time-span based representation, while media and time series files are linked in unanalyzed. This format allows generation of ELAN format files for download to the client tool without regard to the original source form of the annotation file. The database structure further enables the potential for flexible search of the stored annotations both within and across multiple annotation types.

4 The TeraGrid Portal

The large-scale multimedia data collected for multimodal research poses significant computational challenges. Signal processing of gigabytes of media files requires processing horsepower that may strain many local sites, as do approaches such as multi-dimensional scaling for semantic analysis and topic segmentation. To enable users to more effectively exploit this data, the SIDGrid provides a portal to the TeraGrid (Pennington, 2002), the largest distributed cyberinfrastructure for open scientific research, which uses high-speed network connections to link high performance computers and large scale data stores distributed across the United States. While the TeraGrid has been exploited within the astronomy and physics communities, it has been little used by the computational linguistics community.

The SIDGrid portal to the TeraGrid allows the user to specify a set of files in the repository and a program or programs to run on them on the Grid-based resources. Once a program is installed on the Grid, the processing can be distributed automatically

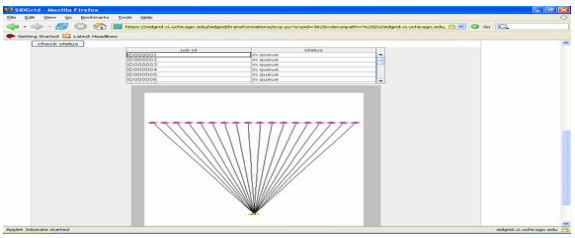

Figure 4: Progress of execution of programs on TeraGrid. Table lists file identifiers and status. Graph shows progress.

to different TeraGrid nodes. Software supports arbitrarily complex workflow specifications, but the current SIDGrid interface provides simple support for high degrees of data-parallel processing, as well as a graphical display indicating the progress of the distributed program execution, as shown in Figure 4. The results are then reintegrated with the original experiments in the on-line repository. Currently installed programs support distributed acoustic analysis using Praat, statistical analysis using R, and matrix computations using Matlab.

5 Prototype Use

The system has been applied to spoken and multimodal discourse and dialogue data ranging from recordings and annotations of multi-party interactions to oral history data to the Talkbank corpus[3], including child language data. This data served as a corpus for basic development of system capabilities. The developers converted the data from their original formats for integration into the repository. The publicly available Talkbank data, such as audio and video media files, can be viewed, browsed, and downloaded from the repository and manipulated in the client-side annotation tool. Prosodic extraction experiments have been performed both using the local client and on the Teragrid, using the dispatch procedures to concurrently analyze data and media files on widely distributed hardware resources. Pitch extraction processes, where analysis of a single file runs out of memory on 2GB, dual-processor Opteron machines, can be completed on 10 files in 1.5 hours with Grid-based servers. These tasks illustrate the scalability of large-scale computation-

[2]http://www.multimodal-annotation.org

[3]http://www.talkbank.org

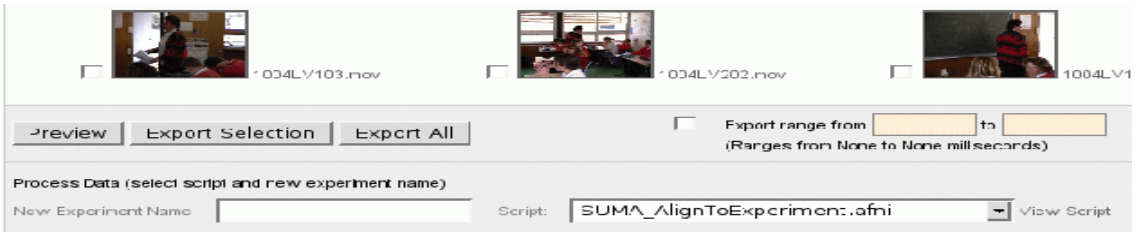

Figure 3: Screenshot of the archive download interface, with thumbnails of available video and download and analysis controls.

ally expensive analyses supported by the SIDGrid framework.

In addition, some preliminary experiments to assess multimodal search and analysis were conducted. These experiments considered the interaction of prosodic features, such as pitch, with other modalities such as gaze or head movement, within turns. These trials demonstrated the capability of search across multiple annotation tiers - including manual speech transcriptions and turn annotations - and time series data from pitch tracking.

6 Future Directions

The SIDGrid infrastructure provides a powerful and flexible environment for annotation, archiving, and analysis of multimodal data. The novel, extensible integration of annotation and analysis both in the client and in the Grid portal will support greater ease of data exploration and large-scale data analysis. The overall framework supports both local data access and distributed annotation and analysis via access to the repository and TeraGrid.

While the basic infrastructure developed thus far is already useful, many extensions to functionality are underway. A major focus is the enhancement of search functionality, for both data and meta-data search. We aim to support both aggregate search for text and annotations across sets of files in the repository under a range of user-specified constraints and search for images in the video recordings. Access to the SIDGrid software and systems is possible through http://sidgrid.ci.uchicago.edu. Future users of the system further will guide its development as new needs come to light.

Acknowledgments We thank other members of the SIDGrid group, including Rick Stevens, David Hanley, Kavithaa Rajavenkateshwaran, and Thomas Uram. This work was supported in part by NSF under Grant No. BCS-05-37849.

References

Jens Allwood, Leif Groenqvist, Elisabeth Ahlsen, and Magnus Gunnarsson. 2001. Annotations and tools for an activity based spoken language corpus. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*, pages 1–10.

S. Bird and M. Liberman. 2001. A formal framework for linguistic annotation. *Speech Communication*, 33(1,2):23–60.

P. Boersma. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5(9–10):341–345.

J. Carletta, S. Evert, U. Heid, J. Kilgour, J. Robertson, and H. Voormann. 2003. The NITE XML Toolkit: flexible annotation for multi-modal language data. *Behavior Research Methods, Instruments, and Computers, special issue on Measuring Behavior*, 35(3):353–363.

M. Kipp. 2001. Anvil- a generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370.

Rob Pennington. 2002. Terascale clusters and the TeraGrid. In *Proceedings for HPC Asia*, pages 407–413. Invited talk.

T. Schmidt. 2004. Transcribing and annotating spoken language with EXMARaLDA. In *Proceedings of the LREC-Workshop on XML-based richly annotated corpora*.

P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. 2006. Elan: a professional framework for multimodality research. In *Proceedings of Language Resources and Evaluation Conference (LREC) 2006*.