DATABASE OF EEG/ERP EXPERIMENTS

Petr Ježek and Roman Mouček

Department of Computer Science and Engineering, University of West Bohemia
Univerzitní 8, 306 14 Pilsen, Czech Republic

Keywords: Electroencephalography (EEG), Event related potentials (ERP), EEG/ERP data, Metadata, Relational database, Web interface, Spring framework, Hibernate, Semantic web technologies, Data conversion.

Abstract: The article deals with the database of EEG/ERP experiments and its developed prototype. Storage, download and interchange of EEG/ERP data and metadata through the web interface is possible, various user roles are defined. The requirements specification including the system context, scope, basic features, data formats and metadata structures is presented. The system architecture, used technologies and the final realization are described. Additional tools and structures as converters of data formats and generated ontology are mentioned. The possible users of the database are specified.

1 INTRODUCTION

Our research group at Department of Computer Sciences and Engineering, University of West Bohemia in cooperation with other partner institutions (e.g. Czech Technical University in Prague, University Hospital in Pilsen, Škoda Auto Inc, ...) specializes in the research of attention, especially attention of drivers and seriously injured people. With regard to our research we widely use the methods of electroencephalography (EEG) and event related potentials (ERP). Within our partner network we are responsible for technical and scientific issues, e.g. EEG/ERP laboratory operation, development of advanced software tools for EEG/ERP research, or analysis and proposal of signal processing methods.

EEG and ERP experiments take usually long time and produce a lot of data. With the increasing number of experiments carried out in our laboratory we had to solve their long-term storage and management. Looking for a suitable data store for EEG/ERP data and metadata we encountered series of problems:

- There is no widely spread and generally used standard for EEG/ERP data files within the community.
- Results (interpretations) of EEG/ERP experiments are usually more important than obtained data (some researchers even declare that experimental data have a low value when they are interpreted).
- There is no reasonable and easily extensible tool for long-term EEG/ERP data (metadata) storage and management (the general practice is to organize data and metadata in common file directories).
- There is no practice to share and interchange data between EEG/ERP laboratories (EEG/ERP data are supposed to be secret or unimportant to share them).

2 SPECIFICATION OF EEG/ERP DATABASE

2.1 System Context

Because of hard manual work with large amount of EEG/ERP data and metadata and in face of difficulties mentioned in Introduction part, we decided to design and implement own software tool suitable for EEG/ERP data and metadata storage and management.

The developed EEG/ERP data store (called simply **the system** in the following text) pursues not only our local research but in general it contributes to advancements in human brain understanding. In addition, we believe that such advanced software

tools increase both the efficiency and the effectiveness of neuroscientific research.

2.2 Requirements Specification

The specification of requirements originated from experience of our laboratory, co-workers from cooperating institutions, books describing principles of EEG/ERP design and data recording (e.g. Luck, 2005) and numerous scientific papers describing specific EEG/ERP experiments. It also corresponds to the effort of International Neuroinformatics Coordinating facility (INCF) (Pelt, 2007) in the field of development and standardization of databases in neuroinformatics.

2.2.1 System Users

The system prototype is dedicated for department users and collaborative partners as well as for a limited group of researchers interested in EEG/ERP research. The system is supposed to be widely tested to guarantee the safety of personal information, availability of EEG/ERP resources and their usability for people interested in this research field.

2.2.2 Project Scope and System Features

EEG/ERP database enables clinicians and various community researchers to store, update and download data and metadata from EEG/ERP experiments. System is developed as a standalone product (integration with the software for EEG/ERP experimental design is not a task of this project). The database access is available through a web interface. We need a web server supporting open source (Java and XML) technologies and a database system, which is able to process huge EEG/ERP data. The system is easily extensible and can serve as an open source.

The system essentially offers the following set of features (the number of accessible features depends on a specific user role):

- User authentication
- Storage, update, and download of EEG/ERP data and metadata
- Storage, update and download of EEG/ERP experimental design (experimental scenarios)
- Storage, update and download of data related to testing subjects

The crucial user requirement is the possibility to add an additional set of metadata required by a specific EEG/ERP experiment. The complete overview of the system features and user roles (use case diagram) is available in (Pergler, 2009).

2.2.3 User Roles

Since the system is thought to be finally open to the whole EEG/ERP community there is necessary to protect EEG/ERP data and metadata, and especially personal data of testing subjects stored in the database from an unauthorized access. Then a restricted user policy is applied and user roles are introduced.

On the basis of activities that a user can perform within the system the following roles are proposed:

- **Anonymous user** has the basic access to the system (it includes essential information available on the system homepage and the possibility to create his/her account by filling the registration form).
- **Reader** has already his/her account in the system and can list through and download experimental data, metadata and scenarios from the system, if they are made public by their owner. Reader cannot download any personal data or store his/her experiments into database.
- **Experimenter** has the same rights as Reader; in addition he/she can insert his/her own experiments (data and metadata including experimental scenarios) and he/she has the full access to them. This user role cannot be assigned automatically, a user with the role reader has to apply for it and the new role must be accepted by supervisor.
- **Supervisor** has an extra privilege to administer user accounts and change their user roles according to the policy.

2.2.4 Data Formats

There exists a variety of data formats for storing EEG/ERP data. The more spread formats and formats used in our laboratory include European Data Format (EDF and EDF+) ("EDF", n.d.), Vision Data Exchange Format (VDEF) ("VDEF", n.d.), Attribute-Relation File Format (ARFF) ("ARFF", n.d.), and KIV format (Kučera, 2008).

European Data Format (EDF) contains an uninterrupted digitized EEG record stored in one file (a header record is followed by data records). The header content has a variable length. It identifies a testing subject and specifies the technical characteristics of recorded EEG signal. The data part contains consecutive fixed-duration epochs of the record. Despite its drawback this data format has

been probably the most hopeful attempt to standardize description of EEG data.

Vision Data Exchange Format (VDEF) is used by the technical equipment in our laboratory. EEG record is divided into three files: a header file, a marker file and a data file. The header file based on Windows INI format describes recorded data and provides a limited set of corresponding metadata as the attribute-value pairs. The marker file contains information about markers (their types and timing) in EEG signal. The data file contains raw EEG data.

Attribute-Relation File Format (ARFF) is used in our laboratory as the interface to WEKA software ("WEKA", n.d.). Data and metadata are stored in one ASCII file consisting of two sections. The header section provides a limited set of metadata and it is followed by the data part.

KIV data format is a modification of simple ASCII format of EEG signal, where metadata (file header in ASCII) are stored in XML file and data from electrodes are stored in separate binary files.

The users' requirement on the system is to accept at least three formats mentioned above. An optional requirement is to provide users with conversion tools between these formats.

Standardization of EEG/ERP data format we are also working on (with INCF support) is out of scope of this article.

2.2.5 Definition of Metadata

The data obtained from EEG/ERP experiments are senseless if they are not supported by more detailed description of testing subjects, experimental scenarios, laboratory equipment etc. Metadata are also necessary for an interpretation of performed experiment and for data search and manipulation. Metadata are organized in several semantic groups:

- Scenario of experiment (name, length, description, …)
- Experimenters and testing persons (given name, surname, contact, experiences, handicaps, …)
- Used hardware (laboratory equipment, type, description, …)
- Actual surrounding conditions (weather, temperature, …)
- Description of raw data (format, sampling frequency, …)

There is important that only a small predefined set of metadata is optional to fill in. In addition, a user with the role experimenter has the right to define his/her own metadata.

2.2.6 System Sustainability

The system purpose is not only to serve as a local managing tool for our EEG/ERP research but to serve as a system, which enables sharing and interchange of data between various research groups. Nowadays EEG and ERP data are provided by diverse groups of not only medical communities but scientists or universities as well. The system is therefore developed as open source accepting INCF recommendations. It will be offered as a free managing tool and source of EEG/ERP data within collection of other neuroinformatics data sources.

2.2.7 System Security

The system database contains personal data, which are necessary for interpretation of experiment or for contact with testing subject. Only experimenter has access to personal data of testing persons who took part in his/her experiment. Collection of personal data and their storage are managed according to law.

2.2.8 System Performance

The system database has to work with long EEG/ERP records (usually tens of megabytes) in reasonable time. The main limiting factor is a user internet connection, not the database performance.

3 SYSTEM DESIGN AND REALIZATION

3.1 System Architecture

The system is based on three layer architecture. This architectonic style is supported by selection of programming tools and technologies. We used Java and XML technologies to ensure a high level of abstraction (system extensibility) as well as a long term existence of the system as open source.

3.1.1 Persistence Layer

Persistence layer uses Hibernate framework. It means that relational database and object – relational mapping are supported. Oracle 11g database server is used to ensure the processing of large data files. ERA model of relational database is available in Figure 1; all tables describing metadata extension are omitted to keep the model understandable.
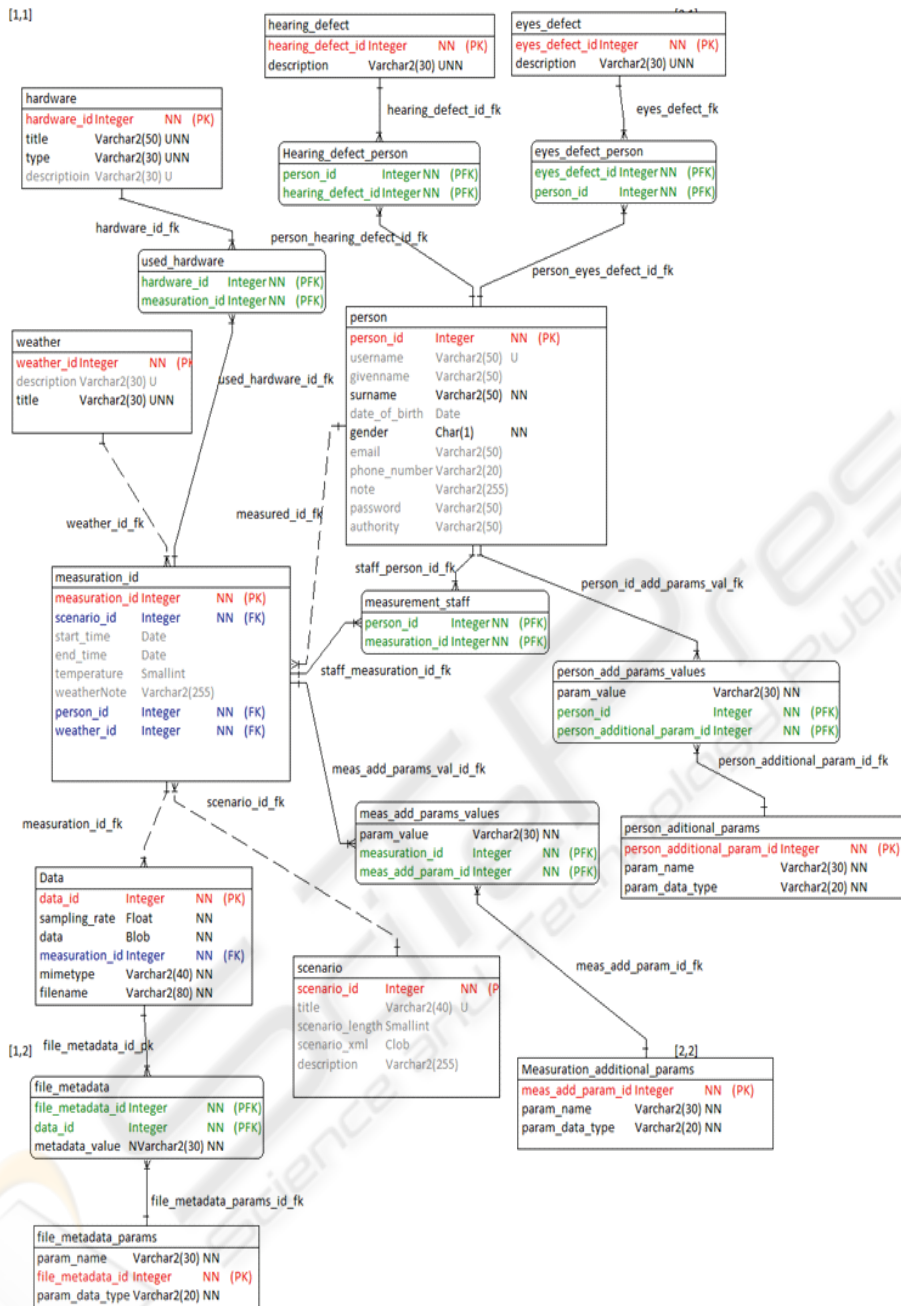
Figure 1: Database structure – ERA model.

3.1.2 Application and Presentation Layer

Application and presentation layers are designed and implemented using Spring technology. This framework supports MVC architecture, Dependency injection and Aspect Oriented Programming. Integration of both frameworks, Hibernate and Spring MVC, was without difficulties. Spring Security framework is used to ensure management of authentication and user roles.

User access to the relational database is realized through the web interface. Majority of users are familiarized with web applications and they do not need any additional software except a web browser.

User interface is divided into several parts (main menu, second level menu, header, footer, and content part). The main menu includes e.g. the

following sections:

- Home – system introduction, registration, login
- Experiments – management of EEG/ERP experiments
- Scenarios – management of experimental EEG/ERP designs
- People – management of people in the system

Figure 2 presents a user interface preview.

Figure 2: User interface preview.

Input data are validated. Error messages are presented using special marks in JSP views and by definition of CSS styles for corresponding input fields.

Storage/download of raw EEG/ERP files is universal; there is possible to store/download any allowed file type.

3.2 Semantic Web Technologies

Registration of the system as a recognized data source occasionally requires providing data and metadata structures in the form of ontology in accordance with ideas of semantic web. We also started to work on the representation of data and metadata structures using semantic web technologies. Nowadays there is possible to generate and provide data and metadata structures using Ontology Web Language (OWL). The details will be presented in a separate paper.

3.3 Conversions Between Data Formats

Converters between data formats mentioned in Section 2.2.4 were implemented. These converters can be downloaded and used locally; no conversion is performed during data upload/download.

4 CONCLUSIONS

The presented system combines research in EEG/ERP and informatics fields as well as application of informatics in neuroscience. Our research group designed and implemented the prototype of experimental EEG/ERP database for storage, download and interchange of EEG/ERP experiments. The database preserves EEG/ERP raw data together with the corresponding metadata. The currently developed prototype is prepared for extensive testing carried out by our department, cooperating institutions and a limited number of people interested in EEG/ERP research and its applications.

Advanced Java technologies (Hibernate, Spring, and Spring security frameworks) were used to ensure a high level of abstraction and further maintenance and extensibility of the system as the open source software.

In addition, converters between various data formats and database ontology in OWL are provided for experienced users.

We hope that EEG/ERP database can also provide useful data and metadata to research groups, which do not perform their own experiments, but which are interested e.g. in signal processing or data mining.

As the next big step we prepare a progressive change of EEG/ERP experimental database to EEG/ERP portal offering e.g. advanced software tools, which can help researchers with difficulties of EEG/ERP experimental design, and set of methods for signal processing.

We also plan to register our system as a data source within large world known projects in neuroinformatics, e. g. Neuroscience Information Framework ("NIF", n. d.).

ACKNOWLEDGEMENTS

REFERENCES

Luck, S. J., 2005. *An Introduction to the Event-Related Potential technique*, MIT Press. Cambridge.

Pelt, J. van, Horn J. van, 2007. Workshop report. *1st INCF Workshop on Sustainability of Neuroscience Databases.* Stockholm.

Pergler, J., 2009. *Database of ERP experiments – business and presentation layer (Databáze ERP experimentů - aplikační a prezentační vrstva).* Thesis (in Czech). University of West Bohemia. Pilsen.

European Data Format (EDF). (n.d.). Retrieved from http://www.edfplus.info/

Vision Data Exchange Format (VDEF). (n.d.). In *Brain Vision Recorder Manual*. Retrieved from http://biomag.uni-muenster.de/EEGlab/docs/VisionRecorderManual.doc

Attribute Relation File Format (ARFF). (n.d.). Retrieved from http://weka.wiki.sourceforge.net/ARFF

Kučera, J., 2008. *Software for ERP processing, data reading and presentation (Software pro zpracování ERP, načítání a zobrazení dat)*. Thesis. (in Czech). University of West Bohemia. Pilsen.

Weka 3 - Data Mining with Open Source Machine Learning Software in Java (WEKA). (n.d.). Retrieved from http://www.cs.waikato.ac.nz/~ml/weka/

Neuroscience Information Framework (NIF). (n. d.). Retrieved from http://www.neuinfo.org/