

Building a Database to Support Intelligent Computational Quality Assurance of Resistance Spot Welding Joints

Lauri Tuovinen, Perttu Laurinen, Heli Koskimäki, Eija Haapalainen and Juha Röning
Department of Electrical and Information Engineering
University of Oulu
Oulu, Finland
Email: {lauri.tuovinen, perttu.laurinen, heli.koskimaki, eija.haapalainen, juha.roning}@ee.oulu.fi

Abstract—A database system for storing information on resistance spot welding processes is outlined. Data stored in the database can be used for computationally estimating the quality of spot welding joints and for adaptively setting up new welding processes in order to ensure consistent high quality. This is achieved by storing current and voltage signals in the database, extracting features out of those signals and using the features as training input for classifier algorithms. Together the database and the associated data mining modules form an adaptive system that improves its performance over time. An entity-relationship model of the application domain is presented and then converted into a concrete database design. Software interfaces for accessing the database are described and the utility of the database and the access interfaces as components of a welding quality assurance system is evaluated. A relational database with tables for storing test sets, welds, signals, features and metadata is found suitable for the purpose. The constructed database has served well as a repository for research data and is ready to be transferred to production use at a manufacturing site.

I. INTRODUCTION

Resistance spot welding is a technique for joining sheets of metal together using strong electrical current. Two electrodes press the sheets tightly against one another and conduct the electricity through them. When the current is high enough, the resistance of the sheets produces an amount of heat sufficient to locally melt the metal. Switching off the current allows the molten metal to resolidify, resulting in a spotlike joint. Employed widely by the automotive and electronics industries, among others, resistance spot welding is an effective and economical welding technique [1].

The bulk of spot welding quality assurance is currently carried out through process control combined with destructive inspection of samples [1]. The problems inherent in such an approach are obvious: a significant portion of produced units goes to waste, and no first-hand information is gained on the quality of the units shipped out. Nondestructive inspection methods exist, but they are not free from problems either, so there is a demand for an inexpensive and reliable way of monitoring welding quality on the production line [1]. An increasingly plausible candidate is computational estimation based on data mining [2], [3], [4], [5], [6], [7]. The key idea is to use signal data recorded during welding to train a classifier,

which is then able to compute a quality estimate for new joints. The extent of destructive testing required can be significantly reduced in this manner. Moreover, data mining can also be used to find near-optimal parameters for a new welding process by analyzing processes recorded in the past. The potential to considerably reduce the cost of producing durable joints is therefore evident.

An essential part of a quality assurance system is a database that holds the recorded data. The database is the component that allows the system to keep all of its accumulated operative intelligence available to appliers. Designing such a database poses many challenges: issues such as intelligent retrieval and flexible extension are critical. In this paper we present a model of a spot welding quality assurance database. A particular point of interest is making communication between the database and other system components as seamless as possible via a set of software interfaces tailored for this purpose. The main contributions of the paper are a schema for constructing the database and a framework for utilizing it using computational methods reported elsewhere to extract weld quality knowledge from the data.

The notion of a welding database has been brought up in several studies ([8], [9], [10]), but no detailed designs appear to have been reported, at least not for systems centered on the idea of classification based on features recorded from welding instances. The novelty of the concept of a data mining-based quality assurance system and the scale of its potential benefits to industry combined with the key role the database plays in the system mean that a good database design is a crucial objective. We believe that the nontrivialities associated with creating such a design justify a devoted paper.

II. DATA MODEL

Let us first consider how the database is populated and utilized, shown in Fig. 1. Case a) depicts the sequence of gathering data into the database through controlled experiments. Case b) shows what happens when a new welding process is to be executed. The welder first records some sample welds, effectively creating a stand-in for the new process. The stand-in is then compared to processes in the database in order to

find the closest equivalent. If an adequate match is found, its optimal parameters can be used to set up the new process; if not, new experiments are required.

The natural fundamental unit of spot welding data is a weld. Associated with each weld are characteristic values such as electrode force and quality measure (e.g. spot diameter), as well as signals (e.g. current and voltage) recorded during the weld. Welds are grouped into sets of units produced as an unbroken series, representative of a particular type of welding process. Exactly what constitutes a distinct welding process is difficult to define formally, so we rely instead on welders to identify representative processes and to provide test sets that make up logical entities.

The raw signal data collected during welding tests is far too noisy and too high in volume to serve as input to a classifier. A number of preprocessing and feature-extraction algorithms therefore need to be executed in order to come up with a practical set of values. It is not strictly necessary to store these values permanently, but in practice there is no reasonable alternative to persistent storage, as it would be very inefficient to recompute the values every time they are needed instead of having them constantly available in a database.

The above analysis yields an entity-relationship (ER) model with four entities and three aggregation-type binary relationships, shown in Fig. 2. Only some of the most important attributes for the 'Test Set' and 'Weld' entities are shown. The exact composition of the attribute sets depends on the needs of the applier, but in any event the number of attributes in a realistic application is probably much higher than in the figure. Also, it was found that new test sets are likely to introduce weld attributes that were not included in the initial specification. Many other attributes were present in the data used in the design of the database and the associated mining algorithms, but they have been omitted from Fig. 2 for the sake of clarity.

The 'Is Preprocessed' attribute of the 'Signal' entity indicates whether the signal is a series of raw measurements or a computationally smoothed-out curve that preserves the overall geometry of the raw signal. However, instead of mapping this attribute to a column in a table it was used to split the 'Signal' entity into two tables: one for raw signals, one for preprocessed ones. Binary large objects (BLOBs) were used to represent vectors of signal values, thus avoiding the creation of a 'Values' table with only one non-key column and easily thousands of rows per signal.

The 'Feature' entity could have been mapped in a fairly straightforward manner, but it was considered more prudent to create a table in which the values of each feature are held in a separate column. This means that the structure of the table needs to be altered when a new feature is introduced, but on the other hand it makes it considerably simpler to manage queries involving large sets of features. The 'Test Set' and 'Weld' entities were mapped straightforwardly, with one additional table to account for the multivalued attribute 'Materials'.

The tables created are summarized in Table I. Each row in the table shows the name of a database table, the name of the

corresponding entity in the ER model, the purpose of the table and a summary of its principal contents. In addition to the data tables derived from the ER model there is a metadata table, 'Column_info', which holds semantic information regarding columns in the data tables. This information is meant to be displayed in the software interfaces through which the database is accessed. The interfaces are described in the next section.

TABLE I
A SUMMARY OF TABLES IN A SPOT WELDING DATABASE

Table	Entity	Purpose	Contents
Test_set	Test Set	Storing attributes of welding test sets	Equipment used, values of constant parameters, quality indicator, threshold of acceptable quality
Material	Test Set	Storing information about materials used in tests	Type and thickness of material, additional notes
Weld	Weld	Storing attributes of individual welds	Values of variable parameters, measured quality, other measured values (e.g. real current level)
Signal	Signal	Storing raw signals	Measured values of current and voltage
Signal_pp	Signal	Storing derived and preprocessed signals	Filtered values of current, voltage, resistance and power
Features	Feature	Storing features	Values of features computed from preprocessed signals
Column_info	n/a	Storing metadata	Column types and descriptions, measurement units where applicable

III. USING THE DATABASE

Fig. 3 shows the database and the operations performed on it by end users. Three software interfaces are required: a data conversion interface for storing raw data, a data export interface for reading stored data and a data import interface for storing refined data. Each interface is a graphical front-end to a query generator that translates instructions specified through dialogs into SQL statements, which are executed by the database management system upon a direct command from the user (in the case of the conversion interface) or upon a request by another software component (in the case of the export and import interfaces). The dynamic nature of the database generates software design issues beyond those of devising appropriate queries and visualizations. The structure of the database is likely to require modifications over time, and this must be accounted for in the access interfaces in order to avoid recompiling. Furthermore, highly transparent interoperability with computational components is necessary for the overall system to operate smoothly. We present our prototype interfaces in this section as an example solution.

The data conversion interface is shown in Fig. 4. The main problem here is that of storing weld attributes: the set of attributes supplied with test sets is not constant, and neither is the set of attributes existing in the database. To this we have devised a solution in which the attributes supplied,

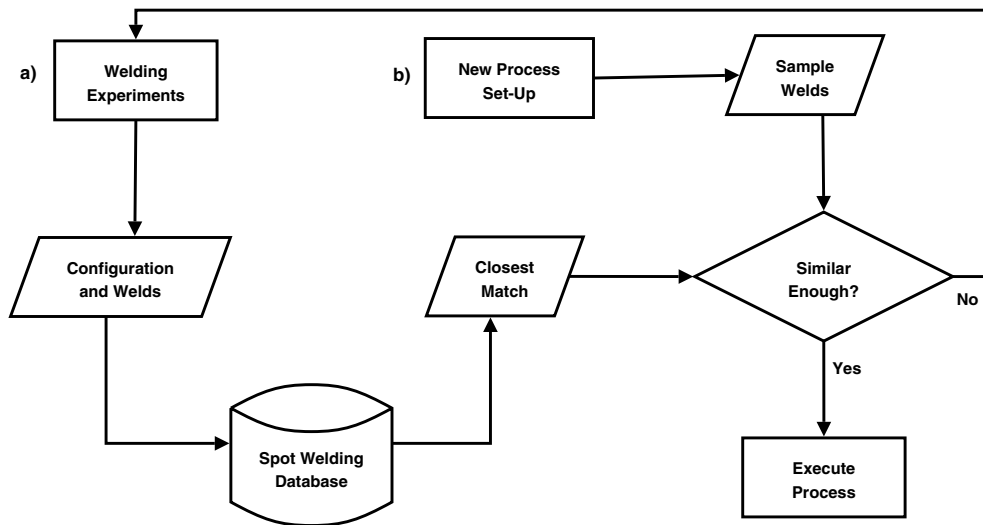


Fig. 1. The spot welding database in its industrial context. A welding engineer at a manufacturing plant accesses the database when setting up a new welding job (b), using a process similarity measure to determine best-guess values for operational parameters. If there is too much uncertainty, the system needs to be adapted to the new process by conducting additional experiments (a).

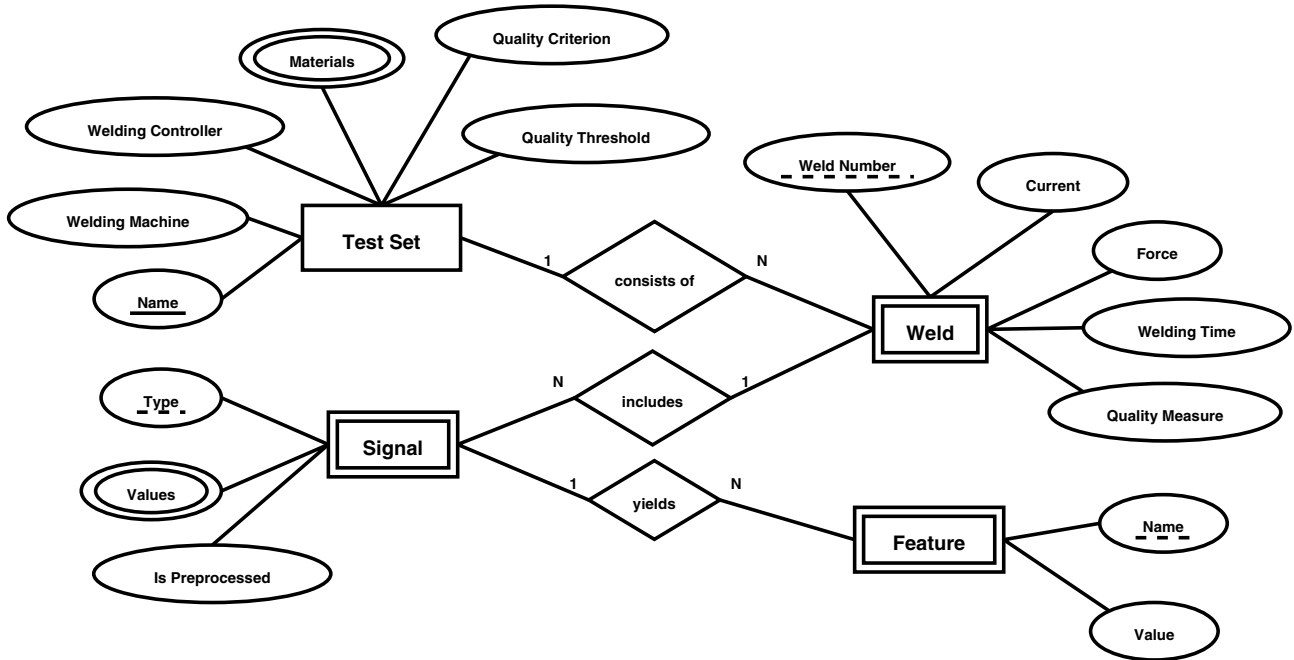


Fig. 2. An entity-relationship (ER) diagram for a spot welding quality assurance database. The ER model identifies the aspects of the application domain to be represented in the database and serves as a starting point for designing the structure (schema) of the database. As a rule of thumb, entities (boxes) become tables in the database and attributes (ellipses) become columns in the tables. Relationships (diamond shapes) are represented by foreign-key columns, which are used to associate rows in a table with rows in another table when a particular relationship exists between them.

represented by columns in a data file, are attached one by one to columns in the database. The list of existing database columns is constantly displayed to the user. If a new attribute is introduced, the user can define a new database column using the appropriate controls without upsetting his or her workflow. Recorded signals are read from a set of binary data files, placed in a directory specified by the user and named such that they can be associated with the corresponding attribute values. The main advantage of this approach is flexibility: the exact format

of the attribute input file is of no import, and an attribute not encountered previously can be accommodated without using a separate database client. The most serious drawback is the amount of somewhat tedious interaction required in mapping input file columns to database columns, although this process could be partially automated.

Fig. 5 shows the data export interface. Before querying the user can view the characteristics of test sets stored in the database and the semantics of available search terms.

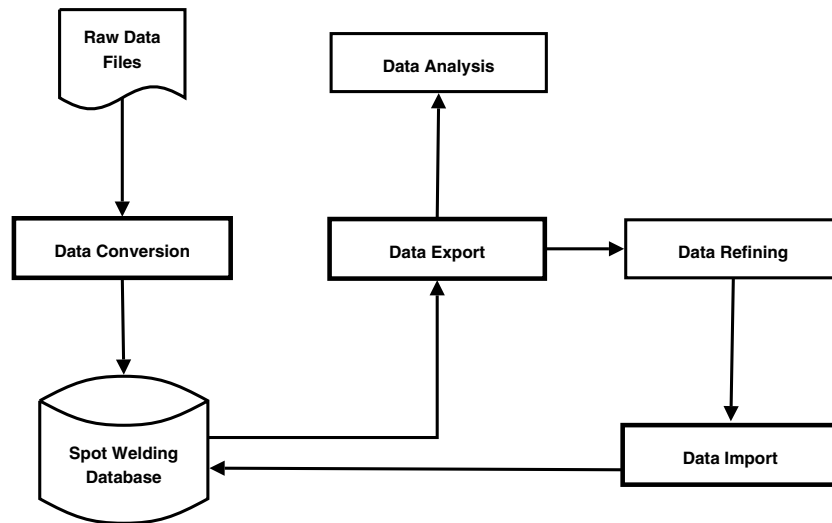


Fig. 3. The data processing operations associated with the database. Each of the interfaces between the database and the real world (shown as boxes with bold outlines) takes on the form of a dialog-based software component. The export and import interfaces are further coupled with other programs to compose processing sequences that produce information required for welding joint quality predictions and welding process similarity computations.

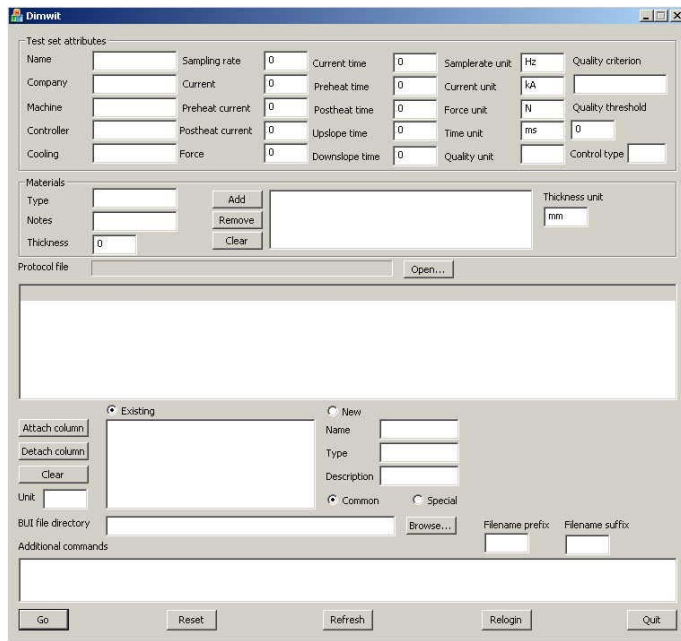


Fig. 4. The data conversion interface, which is used for inserting new test sets into the database. Information regarding an entire test set is input manually into the edit fields at the top. Information on individual welds is contained in a formatted text file, which opens into the panel in the center. Signal data is read from a set of binary files residing in a specified directory. The edit field at the bottom allows the execution of arbitrary SQL commands.

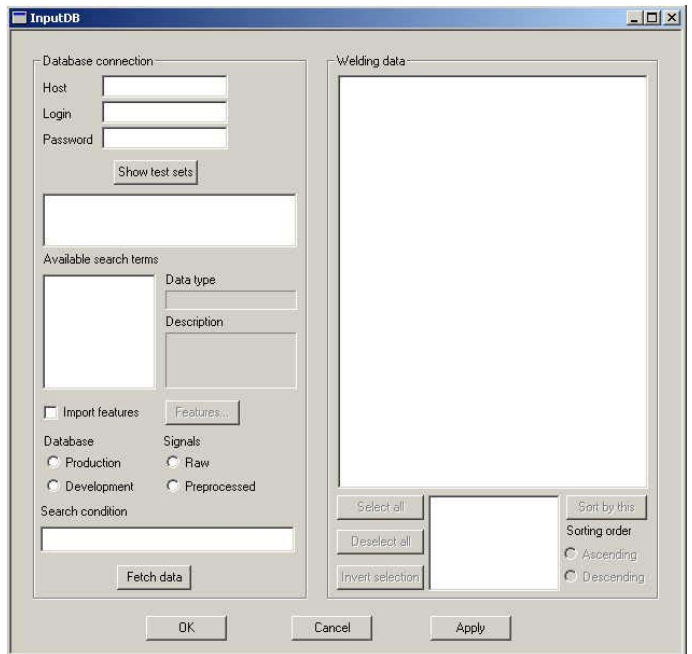


Fig. 5. The data export interface, which is used for retrieving data from the database. The dialog allows the user to browse test sets and examine their properties before formulating a query. The results of the query appear into the list box on the right; selected welds are exported from the database when a computational component requests data from the interface.

The search condition is basically an SQL WHERE clause, another feature designed primarily for flexibility. It allows the construction of arbitrarily complex queries without doing serious harm to usability, since the most commonly needed expressions are soon learned by rote. After querying the user may view attribute values and select the welds to be exported. Finally, the data import interface, fairly simple in comparison,

is shown in Fig. 6. A complication arises only when storing features, in which case the user needs to manually specify the list of features to be imported. This is accomplished through a separate mapping dialog, not shown here. One noteworthy feature is the 'action on duplicate' selection, which allows the user to update an existing test set either by overwriting it or by augmenting it.

Raw welding data is inserted into the database using the



Fig. 6. The data import interface, which is used for writing refined signals and extracted features into the database. Most of the required information is contained in the incoming data packet, so besides database login parameters there is little that the user needs to specify when not storing feature values. Importing features requires additional manual input, supplied through a subdialog opened by clicking on the Select button at the lower right.

conversion interface. It is then accessed via the export interface and refined with various algorithms. The results of these algorithms are written into the database using the import interface. When there is enough refined and classified data in the database it can be used to train a quality prediction model [5], [7]. One can also use a classifier to identify the closest equivalent of a new welding process among those stored in the database; this can be done with high confidence, as a suitable choice of classifier and feature set yields a classification accuracy of over 98% [2], [3]. Process identification is useful for a priori quality assurance, because parameters associated with the best quality in the closest match are likely to result in good quality in the new process as well [4], [6].

The refining and analysis components in the system process data in the form of packets that the components can pass around between them through shared memory. The export and import interfaces were designed to conform to the packet format, so that anything produced by the export interface will be readable by the computational components and anything produced by the computational components will be readable by the import interface. This allows the system section concerned with data preprocessing and mining to interact with the data storage section diversely and efficiently with only a low level of coupling between the two system sections.

IV. EVALUATION

There are many inherent benefits of databases that speak for the adoption of one as an infrastructure for computational quality assurance. A database management system automatically provides a centralized access point through which the data on welding processes is concurrently available to any number of arbitrarily distributed manufacturing locations. Access to the data can be easily restricted in a rich and fine-grained manner. Most important, however, is the hiding of the details of physical storage behind a query engine that allows complex requests to be formulated as concise expressions in a readable, standardized language. This is essential to any data mining

project, which would otherwise get bogged down in such secondary tasks as the design of file formats and access protocols.

The performance of the database was tested with a program that simulated a heavy load on the database by executing repeated retrieval queries from multiple hosts simultaneously. Both the query rate and the level of concurrent access were significantly higher than the system is expected to handle in production use, but response times were nevertheless adequate by a broad margin, indicating that performance requirements would be easily satisfied. The more interesting question is therefore how well the software interfaces to the database serve their purpose.

The most thoroughly tested of the interfaces is the data export interface, which was the first one created. It makes use of the metadata table described in Section II to assist the user in selecting appropriate query terms by displaying information about data fields available for formulating queries. Whether the user wishes to work with raw signals, preprocessed signals or features, the required data is always accessible through the same interface, so switching between tools is never necessary. The export interface has been found well suited to its purpose in all respects.

The data import interface, like the export interface, was designed to be capable of handling both signals and features. The main problem identified in the interface is the necessity of specifying feature names manually one by one when using the interface to write extracted features into the database. Eliminating the problem would have required a means to carry the feature name information in the data packet along with the feature values; however, the format of the data packet was externally specified so this solution was unavailable.

The data conversion interface was created last and is the least tested of the three. Before the interface existed, new test sets were inserted into the database by modifying the source code of a program written for the purpose and then recompiling and executing the program. The deployment of the conversion interface was a considerable improvement over the previous method as it allowed the specifics of new test sets to be encoded parametrically, without any modifications to the source of the conversion routines. The interface also significantly reduced the need of the user storing the new data to be familiar with the structure of the database. Compared to the export and import interfaces the conversion interface is still somewhat complicated and requires more training to operate, but this is to be expected when there has been less time to gather and evaluate user experiences for designing improvements.

With the database and the access interfaces in place, the data mining processes for welding process identification and welding joint quality prediction could be carried out using the truly stepwise approach described in [11]. This means that the results of intermediate transformations are kept in a database so that a mining process may be resumed or restarted at any point between two consecutive major phases. As reported in the reference, such an approach considerably simplifies the

job of a welding expert setting up a computation, which is desirable because the preprocessing sequence alone can be quite complicated. Without a database this breakdown of work cannot be achieved, making computations more error prone and time consuming.

Although [11] is primarily concerned with research work, the results hold for production work as well. This is clear from Fig. 1, which shows that a welding engineer may need to conduct new welding experiments when introducing a welding process of a type not yet represented in the database. In this case the engineer has to write the new test set and the refined information computed from it into persistent storage, a task in which he or she relies on the stepwise data mining process and the entire range of functionality provided by the spot welding database system. In any event, welding process identification requires a comparison of the features of the process to be identified with the features of known processes, and for reasons cited above the only sensible option for making these features accessible is a database.

Experiences with the software interfaces are mainly those of the researchers who used the database as a tool for developing computational quality assurance methods for resistance spot welding. How closely these experiences can be expected to match those of industrial applicators is not clear, but it can already be stated with confidence that the hierarchical database design presented in Section II is logical and comprehensible and thus lends itself well to maintenance and access. The results of the research on quality assurance methods, which relied on the database every step of the way and was considered a success by the industrial stakeholders, testify to the basic soundness of the design. It is therefore justified to propose the design as a basis for industrial production implementations.

V. CONCLUSIONS

In this paper a design for a resistance spot welding database was presented, comprising a schema for the database itself and prototype software interfaces for reading and writing welding data. The database and interfaces were created as a subtask of a larger effort to develop data mining-based computational quality assurance methods for spot welding. By maintaining and mining a database of spot welding processes the need for destructive testing of welding joints can be reduced, which allows the production of high quality welds at lower costs.

The design of the database reflects the hierarchy of welding processes, welds and signals identified in the application domain and the refinement of signals via a preprocessed form into features, which was a major part of the research effort. A metadata table was added for storing information that the software interfaces could display to end users as an aid for constructing appropriate queries. An open-ended design was sought for the interfaces so that they would be able to respond to future requirements that did not arise during the course of the research work. The database proved a valuable research tool and is expected to serve well as a model for industrial implementations.

ACKNOWLEDGMENTS

We wish to thank our colleagues at Hochschule Karlsruhe—Institut für Innovation und Transfer [12], Stanzbiegetechnik GesmbH [13], Harms & Wende GmbH & Co KG [14] and Technax Industrie [15] for providing us with data and feedback that were indispensable for us to conduct the research reported in this paper. The research was carried out with financial support from the Commission of the European Communities, specific RTD program "Competitive and Sustainable Growth," G1ST-CT-2002-50245, "SIOUX (Intelligent System for Dynamic Online Quality Control of Spot Welding Processes for Cross(X)-Sectoral Applications)." The paper does not necessarily reflect the views of the Commission, nor does it in any way anticipate the Commission's future policy in this area. Additional support was received from the Graduate School in Electronics, Telecommunications and Automation (GETA) and from Infotech Oulu Graduate School.

REFERENCES

- [1] TWI, "Resistance spot welding (knowledge summary)," http://www.twi.co.uk/j32k/protected/band_3/kssaw001.html, 2004. Referenced Dec 12, 2006.
- [2] E. Haapalainen, P. Laurinen, H. Junno, L. Tuovinen, and J. Röning, "Methods for classifying spot welding processes: A comparative study of performance," in *Proceedings of the 18th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (IEA-AIE 2005)*, pp. 412–421, 2005.
- [3] E. Haapalainen, P. Laurinen, H. Junno, L. Tuovinen, and J. Röning, "Feature selection for identification of spot welding processes," in *Proceedings of the 3rd International Conference on Informatics in Control, Automation and Robotics (ICINCO 2006)*, pp. 40–46, 2006.
- [4] H. Junno, P. Laurinen, E. Haapalainen, L. Tuovinen, J. Röning, D. Zettel, D. Sampaio, N. Link, and M. Peschl, "Resistance spot welding process identification and initialization based on self-organising maps," in *Proceedings of the 1st International Conference on Informatics in Control, Automation and Robotics (ICINCO 2004)*, pp. 296–299, 2004.
- [5] H. Junno, P. Laurinen, L. Tuovinen, and J. Röning, "Studying the quality of resistance spot welding joints using self-organising maps," in *Proceedings of the Fourth International ICSC Symposium on Engineering of Intelligent Systems (EIS 2004)*, 2004.
- [6] H. Junno, P. Laurinen, E. Haapalainen, L. Tuovinen, and J. Röning, "Resistance spot welding process identification using an extended knn method," in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE 2005)*, pp. 9–12, 2005.
- [7] P. Laurinen, H. Junno, L. Tuovinen, and J. Röning, "Studying the quality of resistance spot welding joints using bayesian networks," in *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2004)*, pp. 705–711, 2004.
- [8] J. Huang, S. Nemat-Nasser, and J. Zarka, "Prediction of fatigue life of metallic structures with welded joints using automatic learning systems," *International Journal of Mechanics and Materials in Design*, vol. 1, no. 3, pp. 255–270, 2004.
- [9] I. S. Kim, Y. J. Jeong, C. W. Lee, and P. K. D. V. Yarlagadda, "Prediction of welding parameters for pipeline welding using an intelligent system," *International Journal of Advanced Manufacturing Technology*, vol. 22, no. 9–10, pp. 713–719, 2003.
- [10] Y.-L. Lee and M.-W. Lu, "Fatigue-reliability analysis of resistance spot welds," in *Proceedings of the Annual Reliability and Maintainability Symposium*, pp. 178–184, 1994.
- [11] P. Laurinen, L. Tuovinen, E. Haapalainen, H. Junno, J. Röning, and D. Zettel, "Managing and implementing the data mining process using a truly stepwise approach," in *Proceedings of the 6th International Baltic Conference on Databases and Information Systems (DB&IS 2004)*, pp. 246–257, 2004.
- [12] Hochschule Karlsruhe website: <http://www.hs-karlsruhe.de/>.
- [13] Stanzbiegetechnik website: <http://www.stanzbiegetechnik.at/>.
- [14] Harms & Wende website: <http://www.harms-wende.de/>.
- [15] Technax website: <http://www.technaxindustrie.com/>.