AN OPEN SOURCE ARCHITECTURE TO COLLECT AND ANALYZE PERFORMANCE MEASUREMENT DATA FROM LOCAL HEALTH AUTHORITIES IN THE TUSCANY REGION

Domenico Cerasuolo, Sabina Nuti
Working Paper n. 02-2009

Domenico Cerasuolo*, Sabina Nuti*,

* Laboratorio Management e Sanità (MeS) - Scuola Superiore Sant'Anna di Pisa

Corresponding Author’s address:
*Laboratorio Management e Sanità - Scuola Superiore Sant'Anna,
Via S. Francesco, 18 56127 PISA – Italy
Tel. (39 +) 050 88 3871/3878/3864
Fax: (39 +) 050 883890
Email: direzionemes@sssup.it
Web site: http://www.meslab.sssup.it/_sito/

Acknowledgments
The authors want to thank the whole Mes Lab team. The usual disclaimers apply

Please quote this way:
Abstract

The aim of this article is to present a hardware and software architecture that shows how to combine different open source technologies in order to realize an efficient system to collect and analyze data from a web survey. The framework presented has been used, in the context of the Performance Evaluation System, to measure the level of satisfaction of the employees of the Tuscany Health Authorities, compared with some specific variables in the internal organizational climate.

Using that framework, about 42,000 employees were given the opportunity to fill in a web questionnaire. The survey has collected 20,294 cases.

JEL Classification: I, H, C

Keywords: web survey, local health authorities, performance evaluation system
# Index

Index .......................................................................................................................... 4

Introduction ............................................................................................................... 5

System Architecture .............................................................................................. 7

CAWI module ......................................................................................................... 8

Structure of the dataset ............................................................................................ 9

Analysis and Reporting ......................................................................................... 11

Advantages of the proposed solution ....................................................................... 14

References ............................................................................................................... 15
**Introduction**

Within any business (or more generally in any organization of people), you must be able to access information, process it, and make it available anywhere. This requirement is particularly important in the always-on world in which users expect to be able to pull up any information required from any location.

The Internet and in particular the http protocol and its extensions, is the most standard, efficacious and efficient solution for sharing, handling and storing information.

This article explores the Linux-Apache-MySQL-PHP, or LAMP, Web development framework and shows how that framework can help you build applications capable of managing the entire process of collection, analysis and production of quality reports on datasets derived from web surveys.

The context in which the system has been designed and implemented is the evaluation of the local health authorities in the Tuscany Region. The evaluation system designed experimentally in 2004 in four pilot areas, subsequently implemented in 2005 in all the local health authorities and finally adopted in 2006 also by the University Hospital Authorities, has become with time a fundamental instrument, which accompanies and supports the government at the regional level.

For the final representation of the results of the Performance Evaluation system, six synthesis dimensions have been identified, able to highlight the fundamental aspects of the performance in a complex system such as the Tuscan one.
The dimensions of the evaluation are

- the evaluation of the population’s health levels
- evaluation of the capacity to follow the regional system’s guidelines
- socio-medical evaluation
- external evaluation
- internal evaluation
- evaluation of operative efficiency and of economic and financial performance

In particular the internal evaluation considers the level of satisfaction of the personnel in the local health authorities.

Each year the personnel in the 16 Tuscan local health authorities are required to answer about 50 questions grouped according to dimensions or areas utilized to measure the different aspects of the climate.

Given the complexity of the survey, owing to the number of people and authorities involved, the questionnaires were distributed using a software solution able to collect the answers to the questions in the questionnaire from the authorities involved in the survey using a simple web browser.

Thanks to the software developed, data can be collected using the Internet as a communication channel. During the month of December 2008 the solution proposed in this article was used to carry out the internal organizational climate survey on the entire population of structured employees of the 16 local health authorities in Tuscany, with a total of about 42,000 people. The survey has collected 20,294 cases.
**System Architecture**

The applications that use the Internet and Internet languages to collect information via the web are described with the acronym CAWI: computer assisted web interview. These instruments are already numerous and well-developed.

The architecture that we propose extends the classic functions of CAWI, created as layer software for data collection, proposing a complete platform for the collection, analysis and reporting of web-derived datasets, integrating the entire process in a single platform.

The proposed system has been developed with open source software and technologies. The Open source technologies used are a web server like apache, a script language such as PHP and a relational database management system extremely suitable for web applications such as MySQL (see figure 1).

*Figure 1 Over-all Architecture of the LAMP application*

The Linux Distribution utilized to manage the server which hosts the application is CentOS, an Enterprise-class Linux Distribution.
derived from sources freely provided to the public by a prominent North American Enterprise Linux vendor.

The architecture is divided into two main modules, the first one collects data in a relational data base through a web form and an authentication mechanism (CAWI module), the second one analyzes data and does the reporting (analysis module). In particular the second module is able to perform a univariate analysis for the variables used in the questionnaire and related graphs.

**CAWI module**

The data are inserted into the relational database after a preliminary check has been made on the quality of the data. This phase of clearing up the data involves the use of client side javascript to check and report to Runtime the errors and/or omissions in relation to the proposed questions. Only when all the coherence controls have been made will the system finally accept the data sent.

Client side control of the data inserted has many advantages.

Communication between the client and the server is reduced to a minimum. Client-server interaction requires the following steps:

1. the browser sends the authentication data (login and password)
2. if the authentication is accepted the server sends the browser the form containing the questions contained in the questionnaire. If the authentication phase is unsuccessful the system sends a message stating that there is an error.
3. When the questionnaire has been fully answered, the client’s data are sent to the server for final registration.

When the form is transferred to the respondent’s client server, all the checks on the fields in the questionnaire are carried out by a javascript function validate(form) transmitted together with the module itself after the authentication step.

The form is able to handle all the objects defined by the html standard.

Data transmission is organized in two parts:

- a main page containing the various fields of the forms, which enable the user to make choices and to insert text
- a secondary page which is referred to by the main page and that does the actual “work” of processing and collecting the data. It is a software component found on the server.

**Structure of the dataset**

Each question present in the questionnaire is univocally associated with a whole number type variable. Each response mode is thus associated with a numerical value contained in the interval \((1,N)\) where \(N\) is the number of possible modes and options for the question in the questionnaire associated with the variable. The whole set of variables thus defined makes up the dataset which will contain the answers to the questions in the questionnaire. The dataset is implemented by a relational table.

Relational databases are currently the commonest type of database. I
The fundamental structure of the relational model is the “relation”, i.e. a two-dimensional table comprising rows (tuples) and columns (attributes). The relations are the entities that are considered to be interesting in the database. Each instance of the ‘entity’ will find a place in a tuple of the relation, while the attributes of the relation will represent the property of the ‘entity’. In our database we have to define a relation called “Questionnaires”, the attributes of which (the variables associated with the questions in the questionnaire) describe the answers to the questionnaires. Each tuple in the "Questionnaires" relation will represent a particular “questionnaire”:

```
questionnaires (1, 2008-11-30, 101, 2, 0, 1, 2, 1,1,2,....)
```

In fact the relation is only the definition of the table structure, i.e. its name and the list of attributes of which it is composed. When it is populated with tuples, we speak of a "relation request". Therefore the previous string represents a “Questionnaires” relation request. A representation of the definition of such a relation could be the following:

```
questionnaires (id,data,azienda,A1,A2,...,An)
```

Where id is the sole indicator associated with the i-th questionnaire, date is the date of compilation of the questionnaire, health authority is the health authority where the respondent A1, A2, … works.., An are the n variables associated with the n questions in the questionnaire. At the end of the survey our dataset will comprise all the questions in the questionnaires compiled and will thus have as many tuples as the questionnaires compiled.
Analysis and Reporting

At the end of the survey the proposed system enables the data to be immediately analyzed and the relative report to be produced.

This phase of the process takes place using an on-line module in which the parameters for “constructing the report” are defined.

In particular:

1. the dataset to be used
2. the conditions for selection on the dataset
3. the heading of the report
4. other personal features (for example the color of the graphs, directions for visualizing the histograms, etc.)

In particular point 1 requires the definition of the parameters for connection to the relational database containing the dataset to be analyzed: the name of the database host, the login and the password of the user enabled to access the database and the name of the database must be specified. This implies that the system can be “utilized” with any dataset constructed using the criteria illustrated.

When the setting phase is over the system is ready to carry out the univariate analysis of the variables contained in the questionnaire.

The analysis can be conducted on all the variables or on a subset of the same. The system allows a very complex report to be produced using the following selected condition syntax: Condition and/or/not/condition where condition is an expression in the form variable =/!= constant. These conditions for selection are applicable to all the variables and enable the dataset to be cut both vertically “projecting the analysis” onto a subset of the variables in the questionnaire and horizontally, “selecting” only the tuples which have certain properties. An example of a condition could be the following: let sex be the
variable containing the information linked to the respondent’s sex (1-male, 2-female) and health authority the variable containing the information relative to the health authority where the respondent works. With the condition

\[(\text{sex}=1) \text{ and } (\text{health authority}=101)\]

We communicate to the software that we wish to make a univariate analysis on the tuples of the dataset corresponding to the questionnaires compiled by the male respondents of health authority 101.

In the following figures we show two snapshots of the system: in the first figure the analysis covers the whole dataset (20,204 questionnaires).

There are two questions in the figure and for each one the system constructs a histogram with the relative frequencies, the average, standard deviation and the table containing the absolute, relative and cumulative frequencies.

In the second figure the system generates for the variable “A1 I am proud to work in this health authority” a graph containing the comparison between the total number of respondents and the respondents from the local health authority 2. Also in this case the synthesis table is generated, containing all the data from the univariate analysis (average, standard deviation, absolute, relative and cumulative frequencies).

Selecting all the variables the system thus generates the entire report which can be exported into pdf format.
An Open Source architecture to collect and analyze performance measurement data from local health authorities in the Tuscany Region
Advantages of the proposed solution

The advantages of the solution proposed are the full computerization of the collection and analysis process, the web as a paradigm of deployment, the extreme flexibility of the system that makes it possible to produce a complete and timely reporting output immediately available to the management, adoption of open, deployed and updated solutions from a community of experts and professionals. With the proposed architecture it is possible to process the data derived from surveys conducted via the Web extremely rapidly and give the health authorities an immediate answer with standardized and detailed reporting. The processing can be shared among several subjects who in parallel can generate the reports for the health authorities using the established criteria: the dataset is centralized; there is no intermediate handling of the data and even non-experts can use the instrument and generate reports.
References


Ginige, A. Murugesan, S., (2001) - Web engineering: an introduction, *Sch. of Comput. & Inf. Technol.*, Univ. of Western Sydney, Campbelltown, NSW


