ICIC Express Letters
Volume **10**, Number **4**, April **2016**

 ISSN 1881-803X
pp. 871–876

ANALYSIS SYSTEM DESIGN FOR KNOWLEDGE VISUALIZATION BASED ON MANUFACTURING DATA

Jin Ma[1], Jong-Suk Ruth Lee[1], Kumwon Cho[1] and Minjae Park[2]

[1]Supercomputing R&D Center
National Institute of Supercomputing and Networking
Korea Institute of Science and Technology Information
No. 245, Daehak-ro, Yuseong-gu, Daejeon 305-806, Korea
{ majin; jsruthlee; ckw }@kisti.re.kr

[2]Engineering Innovation Center
BISTel. Inc.
No. 128, Baumoe-ro, Seocho-gu, Seoul 137-891, Korea
mjpark@bistel.com

Received October 2015; accepted January 2016

Abstract. *In this paper, we propose a data analysis system for knowledge-based visualization. The proposed system incorporates all data manipulation outside of the visualization component, such as data mining, and summarization. We describe the process of concept analysis to visualize the result of analysis based on the data of the manufacturing industry.*

Keywords: Data analysis system, Knowledge visualization, Manufacturing data, XMDR (eXtended Metadata Registry)

1. **Introduction.** Recently, the importance of knowledge visualization [1,2] in presenting a great deal of information to form big data in a limited space effectively has been emphasized. Knowledge visualization improves perception of the field of study, by expressing visual information by appropriate methods in consideration of the purpose and use of the information. In addition, data analysis system has been used in many industrial groups; in particular, we expect an improvement in manufacturing productivity through data analysis. Using analysis systems, manufacturers, for instance, focus strongly on finding the hidden quality factors in processes because yield might be increased or decreased by these factors. Therefore, finding causal factor is very important to improving yields. This research was focused on knowledge visualization [3] and data analysis systems. The core part of the proposed system has three components. First, the data crawler; second, the analysis engine; third, knowledge visualization showing the analysis results to the user. XMDR [4,5] was used as the standard for the metadata in the analysis system. The proposal system solves data heterogeneity using XMDR; visualizes results so that it is possible to obtain new information or insights.

In this paper, an analysis system is proposed on the basis of data from manufacturing; however, in order to provide a more efficient visualization of knowledge, a conceptual flow for analysis is designed.

2. Design of the System.

2.1. Analysis system design.

The proposed system consists of three layers as shown in Figure 1. There exist a service layer, data hub layer, and data & engine layer. The service layer performs the role of manager between the system interface and the lower layers. The data hub layer controls access to the eXtended Metadata Registry (XMDR) and manages integration of data. The data engine layer administrates the XMDR [6-8], Repository and

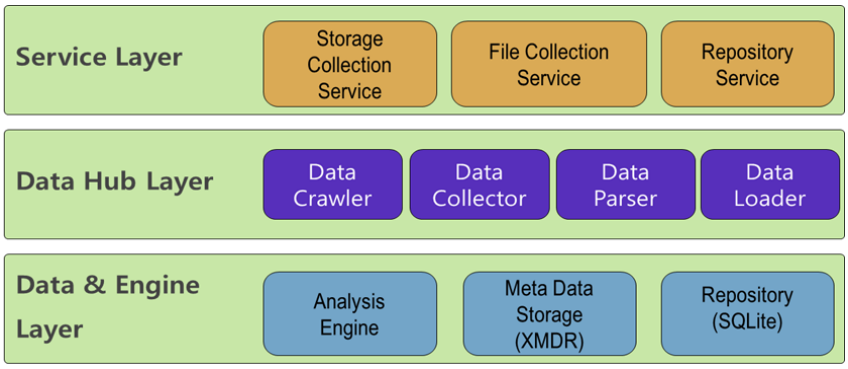

FIGURE 1. Data analysis server architecture

Analysis Engine. The repository is used for saving collected data and creating a table of data. The analysis engine performs data mining and statistical analysis.

2.1.1. *Service layer.* The service layer provides users with service. This layer handles service interface, query conversion, information retrieval, generation of interface. All service modules use the socket communication and XML serialization services.

- Storage Collection Service: A service module that provides an interface between any storage such as RDB, ODB using the XML configuration.
- File Collection Service: File Collection service module that provides an interface between any files such as CSV, TSV, SSV using the XML configuration.
- Repository Service: Repository service module that provides an interface between Repository (SQLite) using the XML configuration.

2.1.2. *Data hub layer.* Data hub layer provides buffer for integrating data. This layer manages data and solves heterogeneity problems. It manages information that was collected from distributed heterogeneous source, and notifies to service layer.

- Data Crawler: Data Crawler to the role of collecting metadata. Collection method uses a set of XML configuration files.
- Data Collector: Data Collector to the role of collecting real data. The Meta information is gathered by crawler using a filter condition, and get data.
- Data Parser: By applying the configuration file, it is to be Parsing to fit the CSV, TSV, SSV file Format.
- Data Loader: When users import the CSV or TSV or SSV file, it is responsible for file loading.

2.1.3. *Data layer.* Data layer takes a role of providing Storage for the proposed system. This layer consists of XMDR, Repository and Analysis Engine.

- XMDR (eXtended Metadata Registry): XMDR consists of MDR and configuration information files. MDR determines the standard of metadata and constructs mapping information by analyzing relationship with metadata of local system. MDR consists of MSO and ML. MSO stores metadata and mapping information for integrating legacy of data layer, ML has location and security information to manage the access of legacy. Configuration file is a repository which has relationship with metadata and data which analyze their collision. Configuration file consists of Database info, crawling and collecting, integration data configuration. ML takes a role of distributed heterogeneous data source mapping information and this configuration file is used for integrating distributed data over the network.
- Repository: Repository is using the SQLite to store the processing result to the user's request. This repository's table structure depends on standard information of XMDR and creates integration tables automatically to perform process.

- Analysis Engine: Plug-in system that provides analytical capability for data processing. (Data Mining, Basic Statistics, Regression, Correlation, ANOVA, Outlier, etc.)

2.2. **Conceptual process of the system.** We describe briefly the analysis system. For ease of understanding, Figure 2 is conceptual process model of proposed system.

Figure 3 shows the analysis system's indexing operation. Different data sources exist at a manufacturing site. A data crawler is an automated service that methodically scans

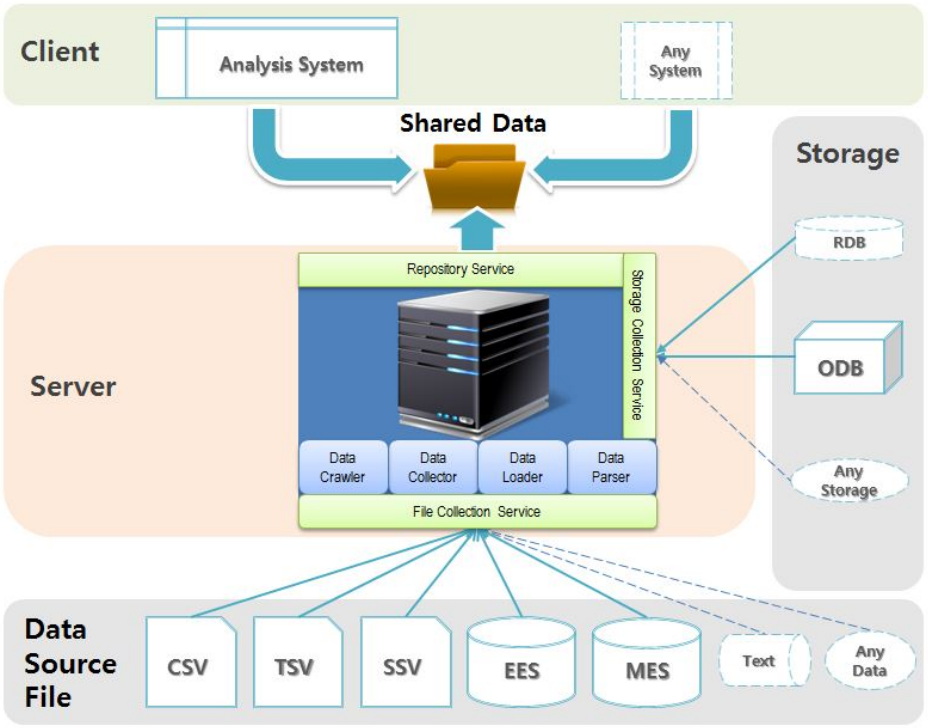

FIGURE 2. Conceptual process model of proposed system

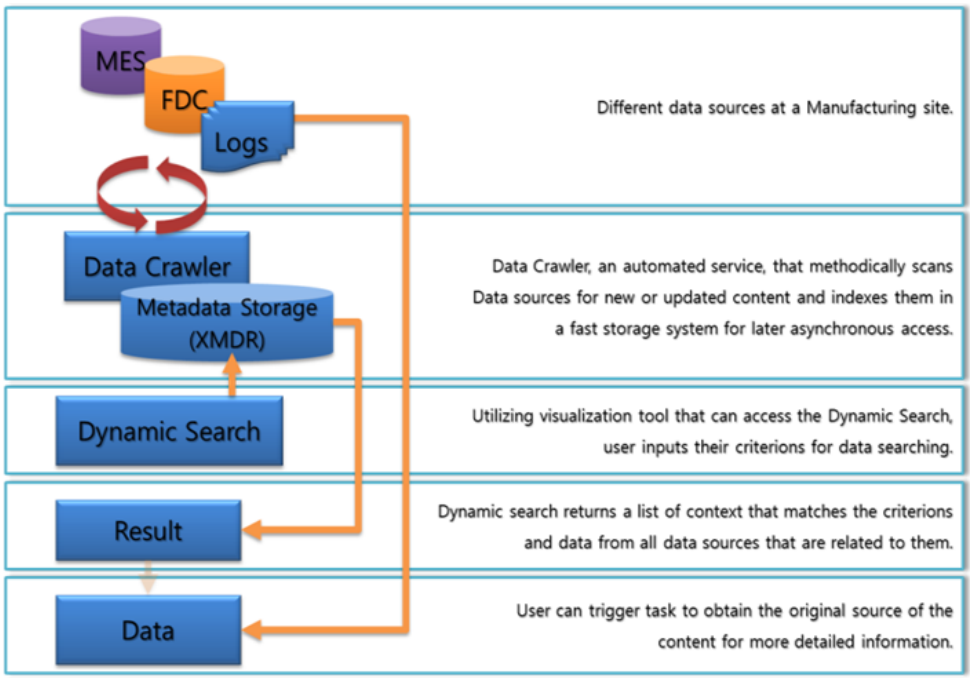

FIGURE 3. Analysis system flow during an indexing operation

data sources for new or updated content and indexes them in a fast storage system for later asynchronous access. Utilizing visualization tools that can be accessed in dynamic search, the user inputs their criteria for data searching. The dynamic search returns a list of contexts that match the criteria and data from all related sources. The user can trigger a task to obtain the original source of the content for more detailed information.

After metadata generation, analysis is performed by the analysis engine. Within the data server component, a module called the analysis engine turns data into meaningful results that can be easily understood by the user (see Figure 4). The role of the analysis engine is as follows:

- Data mining
- Common statistical analysis
  - Correlation
  - Regression
  - ANOVA
  - Summarization (max, min, mean, etc.)
  - Outlier detection

When the analysis is completed the results can be visualized.

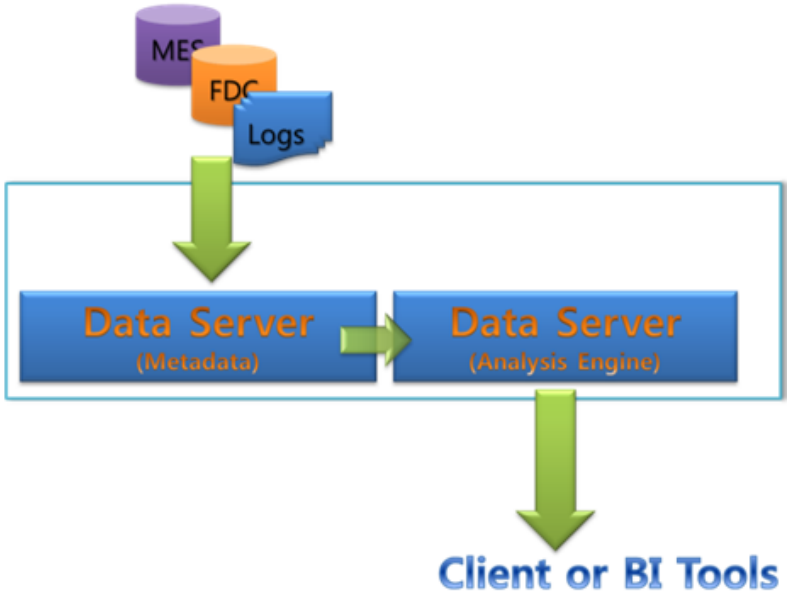

FIGURE 4. Simple process between the analysis engine and metadata

3. **Knowledge Visualization Applications.** Table 1 shows applicable domains and context data sources with use, case example. In order to visualize Chart, Client side developed using .Net with teechart third-party program and Server side is developed using Java. Regression analysis [9-11] is used to infer linear relationships between parameters so that the user can obtain the proper prediction equations for R2R control [12].

By performing an analysis engine regression, we visualize the fitted plot and Q-Q plot. The fitted Plot of Figure 5 is used to display the estimated y variables that are predicted by the regression model. In this plot, the user can view the fitted and actual values at the same time.

The Q-Q plots of Figure 6, the user can check whether or not the residuals of the regression model are approximately normally distributed. Typically, residuals for a good regression model should follow a normal distribution. In the Q-Q plot, residuals are plotted against the reference line on the chart. If the distribution of residuals matches the theoretical distribution, data points form a linear pattern.

TABLE 1. Examples of analysis system indexing

Analysis Domain	Semiconductor	Solar
Use case example	•Yield Loss analysis •Tool control •Recipe management •Inventory control	•Efficiency analysis •Tool control •Inventory control
Contexts	Lot and Wafer	Sheet and Panel
Data Sources	•Tracking •Metrology •Defect •Recipe •Process (Trace) •Alarm •Tool maintenance •WET •Sort Test •Packaging •Final Test	•Tracking •CD •Recipe •Process (Trace) •Defect •IV •Packaging

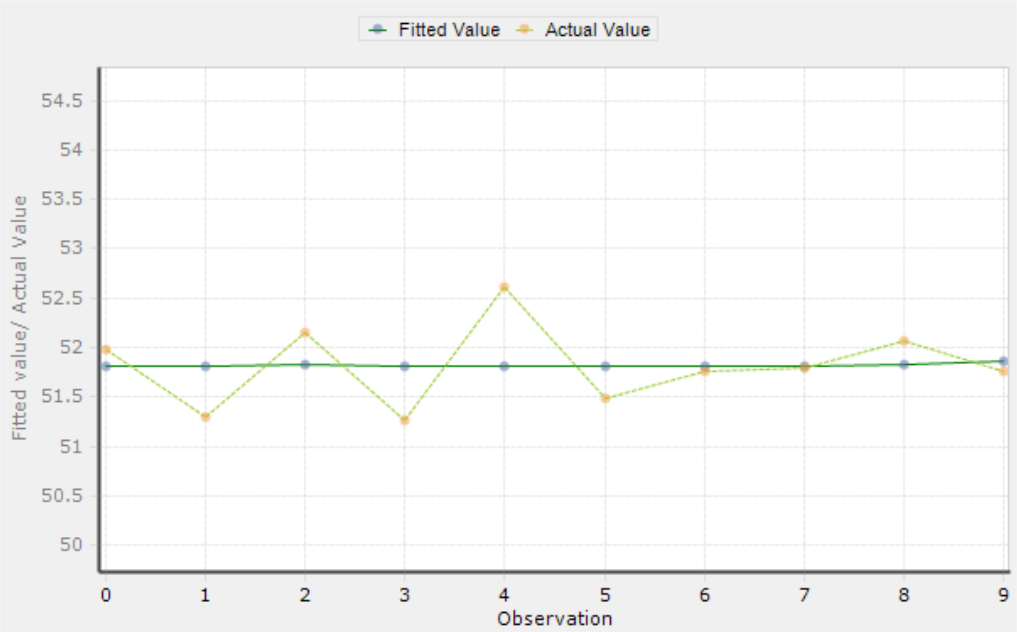

FIGURE 5. Regression visualization: fitted plot

4. **Conclusions.** We have applied a data analysis system for knowledge-based visualization. The analysis system can be applied in manufacturing as well as in other areas. Using the analysis system can find the cause of the defects in processes and increase productivity.

In the manufacturing industry, the proposed system unifies and normalizes all outputs and configurations, reducing the learning curve and enhancing productivity. The proposed system's client program is limited to visualization. Therefore, future work will connect the proposed system with business intelligence (BI) tools. And we will have attempts more advancing and improving performance of data-integration process.

Acknowledgment. This research was supported by the EDucation-research Integration through Simulation On the Net (EDISON) program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. NRF-2011-0020576).

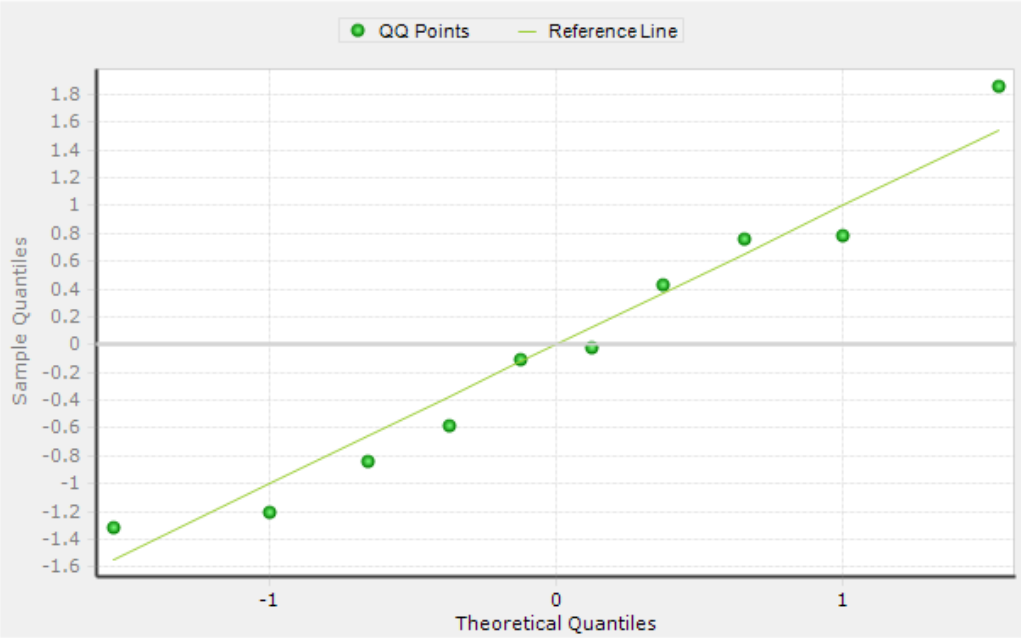

FIGURE 6. Regression visualization: Q-Q plot

REFERENCES

[1] S.-O. Tergan, Digital concept maps for managing knowledge and information, *Knowledge and Information Visualization*, pp.185-204, 2005.

[2] R. A. Burkard, Towards a framework and a model for knowledge visualization: Synergies between information and knowledge visualization, *Knowledge and Information Visualization*, pp.238-255, 2005.

[3] Y. Zhang, X. He, J. Xie and Z. Wang, Study on the knowledge visualization and creation supported kmap platform, *International Workshop on Knowledge Discovery and Data Mining*, pp.154-159, 2008.

[4] ISO/IEC IS 11179, *Information Technology-Specification and Standardization of Data Elements*, 2003.

[5] K. D. Keck and J. L. McCarthy, *XMDR: Proposed Prototype Architecture Version 1.01*, http://www.xmdr.org/, 2005.

[6] J. Ma, S.-J. Moon, G.-D. Jung and Y.-K. Choi, Design and implementation of XMDR based on OGSA-DAI system for data integration retrieval, *Proc. of KIPS Fall Conference*, vol.16, no.2, pp.173-174, 2009.

[7] J. Ma, S.-J. Moon, G.-D. Jung and Y.-K. Choi, Design of multi-agent system based on the P2P networks using query rewriting, *The 34th KIPS Fall Conference*, vol.17, no.2, pp.1780-1783, 2010.

[8] J. Ma, I. H. Jeon and Y. K. Choi, Design of an efficient keyword-based retrieval system using concept lattice, *Journal of Internet Computing and Services*, vol.16, no.3, pp.43-57, 2015.

[9] Regression analysis, *Encyclopedia of Mathematics*, Springer, 2001.

[10] E. J. Williams and I. Regression, Linear hypotheses, in *International Encyclopedia of Statistics*, W. H. Kruskal and J. M. Tanur (eds.), Free Press, 1978.

[11] D. V. Lindley, Regression and correlation analysis, *The New Palgrave: Dictionary of Economics*, 1987.

[12] K. Wang and F. Tsung, Run-to-run process adjustment using categorical observations, *Journal of Quality Technology*, vol.39, no.4, 2007.