A Knowledge Dashboard for Manufacturing Industries

S. Mazumdar[1], A. Varga[2], V. Lanfranchi[2] and F. Ciravegna[2]

[1]Information School
[2]OAK Group, Department of Computer Science
University of Sheffield, Regent Court – 211 Portobello Street,
S1 4DP Sheffield, UK
s.mazumdar@sheffield.ac.uk
{a.varga, v.lanfranchi, f.ciravegna}@dcs.shef.ac.uk

Abstract. The manufacturing industry offers a huge range of opportunities and challenges for exploiting semantic web technologies. Collating heterogeneous data into semantic knowledge repositories can provide immense benefits to companies, however the power of such knowledge can only be realised if end users are provided visual means to explore and analyse their datasets in a flexible and efficient way. This paper presents a high level approach to unify, structure and visualise document collections using semantic web and information extraction technologies.

Keywords: Semantic Web, Information Visualisation, User Interaction.

1 Introduction

Modern manufacturing is a complex domain where productivity and efficiency are strongly affected by a broad range of factors such as site locations, cultural values, management decisions and communication capabilities. For example, large manufacturing organizations are usually globalised, with facilities geographically distributed, making use of multiple manufacturing machines, interacting with several suppliers and warehouses. Also, a recent trend in large organisations has been the presence of dynamic, interdisciplinary working groups and communities of practice who require rapid, flexible customisation of information to their specific needs [1]. At the same time, the information they generate needs to be shared with the rest of the organisation, and hence, must be presented to other communities in ways that can be easily understood (and correctly interpreted) and reused [2].

The underlying commonality between these phenomena is information availability: if information is captured, stored and shared between different departments and locations then efficient communication can be reached and stronger support for managerial decisions can be provided. Unfortunately this information is often collected in a wide variety of formats (e.g., text files, images, PDF documents) and dispersed in independent repositories, including shared directories, local and company-wide databases, ad hoc information systems, etc. Critical knowledge may be hidden in the huge amount of manufacturing data, and the cost of exhaustively identifying, retrieving and reusing information across this fragmentation is very high and often a near impossible task.

This paper presents how Semantic Web and Information Extraction (IE) technologies can be adopted to unify such collections of documents and formalize their knowledge content, bringing together information from different domains, which can feed into organisational knowledge. Visualisation techniques can then be applied on top of the semantically structured data to explore, contextualise and aggregate it, offering multiple perspectives on the information space and provide analytic tools that could support users in spotting trends and identifying patterns and relationships. In order to achieve this goal two steps are required:

- Knowledge Acquisition: acquiring information from different documents and corpora and semantically structuring it in a semi-supervised manner.
- Knowledge Visualisation: creating multiple views over the semantic knowledge space.

Our methodology is innovative compared to previous literature (analysed in Section 2) as it defines the Knowledge Acquisition and Visualisation steps at an abstract level: the use of ontologies to extract, structure and visualise information make our approach flexible, reusable and extensible.

The Knowledge Acquisition and Visualisation steps will be described in details in Section 3, before providing implementation details (Section 4) and discussing conclusions and future work (Section 5).

The following scenario (taken from SAMULET[1], an existing research project on advanced manufacturing in the aerospace industry in which the authors are involved) has been considered as a foundation for the work: in a manufacturing industry a huge number of components are produced every day based on design data provided by Design departments, and are reused in other divisions of the company. When these components are produced manufacturing data is collected such as manufacturing time, location of the plant and of the manufacturing machine, type of component and details (possibly linked to design data). Additional information includes the person and machine responsible for the production, manufacturing costs and so on. This data is collected in a wide variety of formats (e.g. Excel spreadsheets, images, Word Documents), stored in independent repositories and often distributed using personal channels (such as e-mails, or shared network drives).

Manufacturing data are essential to resolving any issue that may arise on a component, in order to be able to clearly identify the driving factors behind the issue and to discover any significant trends or patterns related to individual manufacturing units/machines/personnel. Identifying non-obvious patterns in the data is fundamental to increasing productivity and efficiency: for example, a consistently poorly performing machine may be over-shadowed by a well performing manufacturing unit

[1] SAMULET (Strategic Affordable Manufacturing in the UK with Leading Environmental Technology), http://www.rolls-royce.com/investors/news/2009/280709_research_factories.jsp Last Accessed 14/04/2011

– data analysis and visualisation would help in spotting such trends and support putting corrective measures in place.

2 Related work

Our approach aims to provide a consistent and coherent environment for knowledge exploration in the manufacturing domain, encompassing knowledge acquisition and knowledge visualisation techniques. Related work in both these areas is now analysed, with particular emphasis on the adoption in the manufacturing domain.

2.1 Knowledge Acquisition

Traditional machine learning (ML) approaches for knowledge acquisition in manufacturing started to gain much attention only in recent years [3-10], mostly because the majority of the ML algorithms and tools require skilled individuals to understand the output of ML process [3]. However there has been some work on using traditional ML techniques for specific areas (such as fault detection, quality control, maintenance, engineering design, etc.) employing classification [6,7], clustering [8] and association rule mining [9,10] algorithms [3-5]. Classification algorithms were used for categorising data into different classes, for example classifying defects in the semi-conductor industry [5]. [6] employed a hybrid approach combining neural networks and decision tree classification algorithms for recognising false classifications in control chart pattern recognition (CCPR) thus facilitating quality control. [7] used decision tree algorithms for producing classification rules which were then saved in the competitive decision selector (CDS) knowledge bases enabling efficient job shop scheduling. Clustering algorithms were also used to group similar data into clusters, for example clustering the orders into batches for speeding up the product movement within a warehouse [5]. [8] applied fuzzy c-means clustering algorithm for identifying changes in traffic states thus improving the traffic management systems. Association rule mining algorithms were used to identify relationships among the attributes describing the data. [9] used association rule mining for detecting the source of assembly faults, thus improving the quality of assembly operations. [10] extracted association rules from historical product data to identify the limitations of the manufacturing processes. This information can then be used to improve the quality of the product and identify the requirements for design change.

Despite the increased interest, most of these approaches still lack portability and require a large amount of annotated data to achieve high performance, which is usually tedious and costly [13] to obtain. Furthermore recent advances in domain adaptation show that traditional Machine Learning (ML) approaches for IE are no longer the best choices [11,12]. These algorithms work only well when the format, writing style in which the data (e.g. manufacturing time, location of the plant and the machine) is presented is similar across different corpora [11,12]. In dynamic and heterogeneous corpora, these ML based systems need to be rebuilt for each corpus or format, making them impractical in many scenarios [11], such as the one presented in this paper. To enable effective knowledge capture in manufacturing our approach employs an adaptable IE framework based on domain adaptation techniques, as presented in Section 3.

2.2 Knowledge Visualisation

Information visualisation techniques have been extensively adopted in the manufacturing domain to display and illustrate different processes such as simulation of model verification and validation, planning, decision making purposes and so on [14, 20]. Though most simulation results are based on data models, visualisations are essential to efficiently communicate information to end-users [15]. For example visualising CAD (Computer Aided Design) models enriched with performance scores provides analysts insights into the performances of different manufacturing units; alternative techniques provide ways for manufacturing units to validate their products against software models [14] (to evaluate compliance of manufacturing units to design).

Commercial tools generally focus on 3D visualisations of manufacturing models, factories, machines and so on. Examples of such commercially available tools used in the manufacturing industry include Rockwell's FactoryTalk[2] (remote monitoring of manufacturing processes); Autodesk's 3ds Max[3] and Maya[4] (modelling of product designs, animation, virtual environments); VSG's OpenInventor[5] (3D Graphics toolkit for developing interactive applications); DeskArtes ViewExpert[6] (viewing, verifying, measuring CAD data); Oracle's AutoVue[7] (Collaboration tool to annotate 3D or 2D models). These 3D commercial tools are also adopted in other industries like gaming, animation and so on [17]. However the high cost of 3D hardware and software makes this option unfeasible for smaller companies [16].

3D visualisation techniques have also been investigated in academic works, such as Cyberbikes, a tool for interaction with and exploration using head-mounted displays. [21] presents another example of 3D visualisation, providing factory floor maps which use animations to convey real-time events.

Using visualisations to communicate high-quality data in manufacturing scenarios can greatly reduce the amount of time and effort taken by engineers to resolve an issue: in a study by [18], engineers provided with animated visualisations combining several steps of a simulation could substantially reduce their analysis time. [22] discusses how factory map visualisation based navigation can often provide means to significantly reduce the cognitive load on analysts monitoring a typical manufacturing factory, when compared to list-based navigation of factory machines and their performances. Our approach takes inspiration from this latter works in aiming to

---

[2] FactoryTalk, http://www.rockwellautomation.com/rockwellsoftware/factorytalk/ Last Accessed 14/04/2011

[3] AutoDesk 3ds Max, http://usa.autodesk.com/3ds-max/ Last Accessed 14/04/2011

[4] AutoDesk Maya, http://usa.autodesk.com/maya/ Last Accessed 14/03/2011

[5] VSG OpenInventor, http://www.vsg3d.com/open-inventor/sdk Last Accessed 14/04/2011

[6] DeskArtes ViewExpert, http://www.deskartes.com/ Last Accessed 14/04/2011

[7] Oracle AutoVue, http://www.oracle.com/us/products/applications/autoVue/index.html Last Accessed 14/04/2011

provide efficient visualisation techniques that will reduce engineers cognitive workload and facilitate knowledge analysis.

3. Adding semantics to the manufacturing domain

Given the large scale and the heterogeneity both in data types and data formats, automatic techniques are required to process the data, unifying the document collections and formalising their knowledge content. In the following we distinguish between data, information and knowledge as proposed in [27]. Namely, data refers to the basic raw unit without any implicit meaning, information refers to data enhanced with context and perspective, and knowledge is information connected by patterns and relations. In our case the outcome of our Information Extraction framework is considered knowledge as it extracts entities and relations and assigns semantic meaning to them.

Our approach (shown in Figure 1) is therefore based on the use of a common knowledge representation in the form of ontologies describing the manufacturing domain. The ontologies are created manually so that the high-level ontology covers the generic manufacturing scope (common concepts and relationships between them), and the local ontologies (interlinked by the over-arching high-level ontology) capture the information specific to the different corpora. An adaptable Information Extraction framework considering the high-level ontology then extracts the common concepts across the corpora, thus avoiding ontology mapping and integration (see Section 3.1).

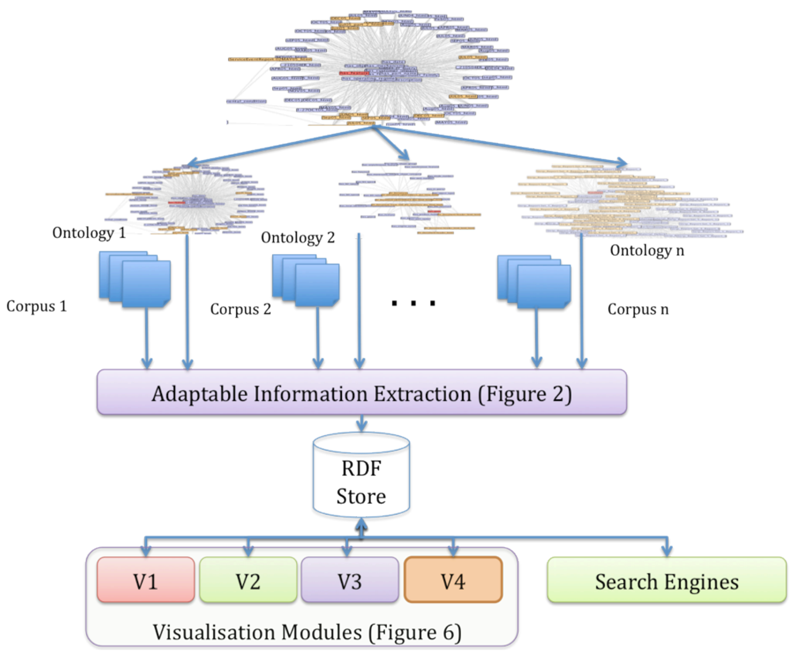

Figure 1 - The knowledge acquisition and visualisation process

The extracted information is then stored in a RDF store and available for query and visualisation (see Section 3.2).

3.1 Adaptable Information Extraction framework

The adaptable Information Extraction (IE) framework runs in a semi-supervised manner over the (automatically converted) textual versions of the documents in each corpus, extracting the relevant entities and relations and mapping them to the ontological concepts. The IE process is composed of two steps:

- Manual annotation of a subset of data by domain experts for training purposes.
- Unsupervised domain adaptation and annotation of the remaining documents using a Support Vector Machine (SVM)[8] [25] classifier.

Whilst this approach is common in literature [11,12] the novelty is in the portability of the classifier between different corpora with minimal supervision (using only a small amount of human annotations). For each new corpus (and document type) the initial classifier is augmented applying a feature representation approach [11,12] inspired on [29]. That is, the words from all the corpora are first clustered into semantic topics using Latent Dirichlet Allocation [28] topic model. Then new semantic features consisting of a set of most probable topics for each word are added to the classifier. This approach makes our IE system flexible and adaptable, enabling efficient knowledge acquisition across corpora.

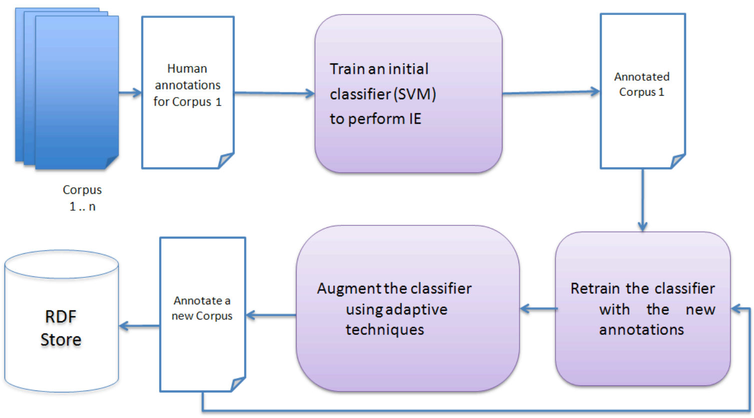

Figure 2 - The adaptable information extraction process

The extracted knowledge (ontology-based annotations) is then stored in a triple store in the form of RDF triples and used later for semantic visualisation. The current implementation of the IE framework also applies a terminology recognition [26] module for domain specific information extraction (e.g. type of component) within the SAMULET project, however the scope of the IE system is more generic and

[8] LibSVM tool: http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/ Last Accessed 14/04/2011

allows extracting domain independent entities and relations too (e.g. person, time, location).

3.2. Knowledge Dashboard

Our approach focuses on providing multiple knowledge visualisations at different granularity levels, using a semantic knowledge dashboard to support users in quickly gathering a broad insight of their datasets from differing perspectives. This approach is based on a set of interlinked ontologies (as explained in Section 3), which structure the knowledge from the different corpora and define relations between found entities. For each semantic entity type in the knowledge space a set of possible visualisations is defined: automatic inferences are then made on the type of entities and relations stored in the knowledge space to create the visualisation widgets. These visualisations can be customised to suit the user task, needs and preferences.

A dashboard interaction paradigm has been chosen as it provides large amounts of information in one interface, without compromising on clarity [23] and it is an increasingly common visualisation paradigm thanks to its adoption by several well-known websites like igoogle[9] and BBC[10]. Such an approach offers the possibility of dynamically choosing the best visualisation tool for the task in hand, as differently represented data can reveal different insights.

A detailed scenario is now presented to highlight the features of the knowledge dashboard and the interaction possibilities. In our hypothetical scenario, a manufacturing engineer (Bruce) working at a large aerospace organisation has access to six types of documents from different departments:

- Machine Performance Reports - describing operational performances of machines at manufacturing sites;
- Site Performance Reports - describing the overall performance of manufacturing sites;
- People Pages – websites of various individuals and authors of the reports;
- Machine Testing Reports – describing the findings of laboratory testing on machines at manufacturing sites;
- Quality Documents – reports discussing the outcome of various quality tests on manufactured products.
- Service Event Reports – reports discussing various service and maintenance operations conducted on engines over their lifetime.

These different report types have been analysed using our adaptable IE framework and semantic knowledge has been extracted and stored in a unified knowledge base. Visualisation ontologies have been defined for the different entities and relations and for the user preferences. These ontologies are used by the knowledge dashboard to automatically build the knowledge space and visualisation widgets. The selections of the visualisations are based on various features such as user preferences, usage history, current task, scale of retrieved datasets and types of data. The visualisation ontologies are essentially classifications of existing visualisations based on these

[9] iGoogle interface, http://www.google.com/ig, Last Accessed 04/03/2011
[10] BBC interface, http://www.bbc.co.uk/, Last Accessed 04/03/2011

parameters. Once a dataset is retrieved, the visualisation ontologies are used to infer the most effective visualisations for the dataset and users.

In our scenario, Bruce is investigating a condition where a lot of enquiries have been made to the manufacturing teams while service engineers were inspecting compressors of several engines during maintenance. Bruce first selects the relevant document sets from the combo boxes provided in the query interface. He then selects the filters 'Regime' and 'Component' and enters his query ('maintenance' and 'compressor' respectively).

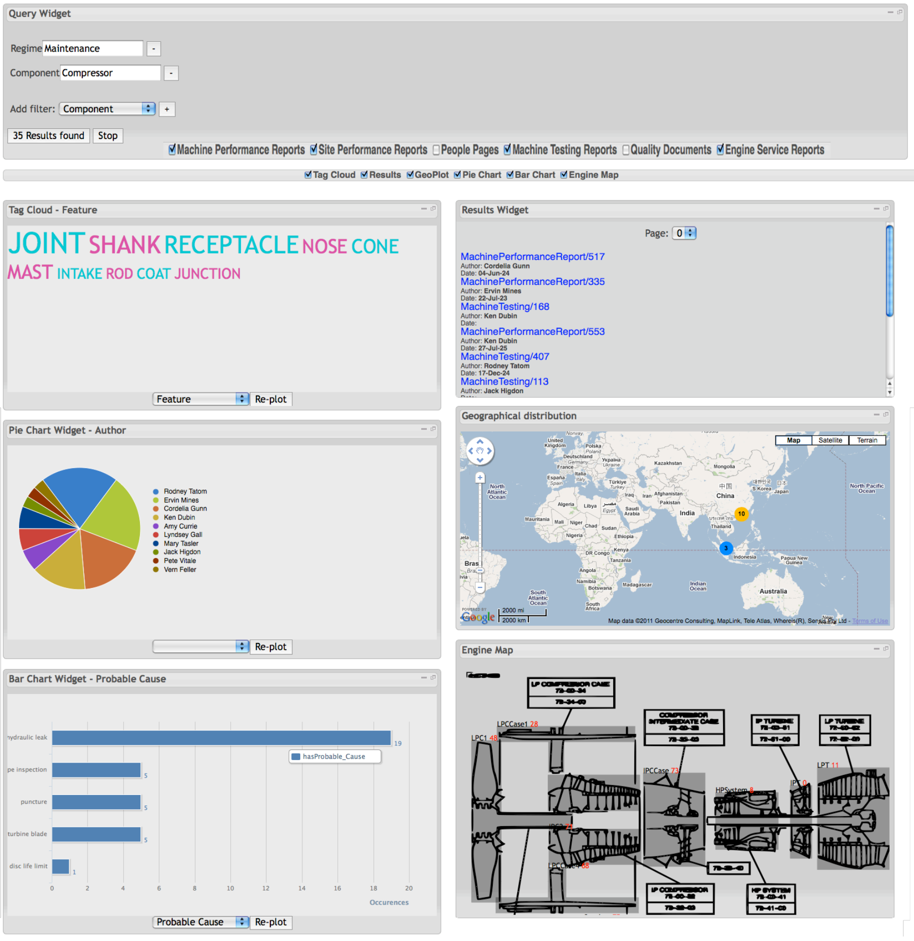

Figure 3 - Knowledge Dashboard

This initiates several queries to be sent to the backend from the interface. Bruce is then provided with several widgets (Figure 3), each of which present different facets (powered by different relations in the underlying semantic knowledge): the tag cloud informs Bruce that out of all the documents retrieved, most of the discussion has been related to features 'joint', 'Shank' and 'Receptacle' – this could indicate which manufacturing machines might be responsible. The bar chart indicates that most of the documents discussing engine service events have also discussed ‘hydraulic leaks’. The engine map groups the documents by the components they discuss – this shows how documents discussing ‘compressors’ also refer to other related components. These components are then displayed as grey areas, along with counts of how many documents have been found for each component. The pie chart provides a plotting by document authors – this enables Bruce to contact authors for further information and advice. The geographical plot provides the locations of manufacturing sites that are responsible for producing the components being described in the datasets. Using such visualisations, Bruce can now answer several common questions often asked during investigations: Where are the manufacturing machines located? What parts of an engine have the machines manufactured? What are the features of the parts that are being manufactured? Who are responsible for the manufacturing sites? From the multiple visualisation layers a summation of the knowledge emerges that can highlight previously unseen trends, patterns and issues/relations.

Thanks to the semantic knowledge and the background ontologies, the document collections can be visualised at different levels of granularity. For example an encompassing visualisation is achieved by displaying the whole document collections and comparing them, to show a high level view on the available facets without having to look at the individual document instances. The widgets are interactive, allowing zooming and selecting the preferred granularity level, from document to instance level. This follows the well-known principle of “overview first, zoom and filter, then details-on-demand” [24].

In our scenario if Bruce needs to analyse the performances of the organization from a manufacturing unit point-of-view, he can explore the knowledge space using a geographic view, then zooming in on an individual manufacturing site to reveal the site’s floor plan along with the positions of the manufacturing machines. This floor plan is then enriched with performance statistics of the machines, extracted from the Site and Machine performance reports, as shown in Figure 4.

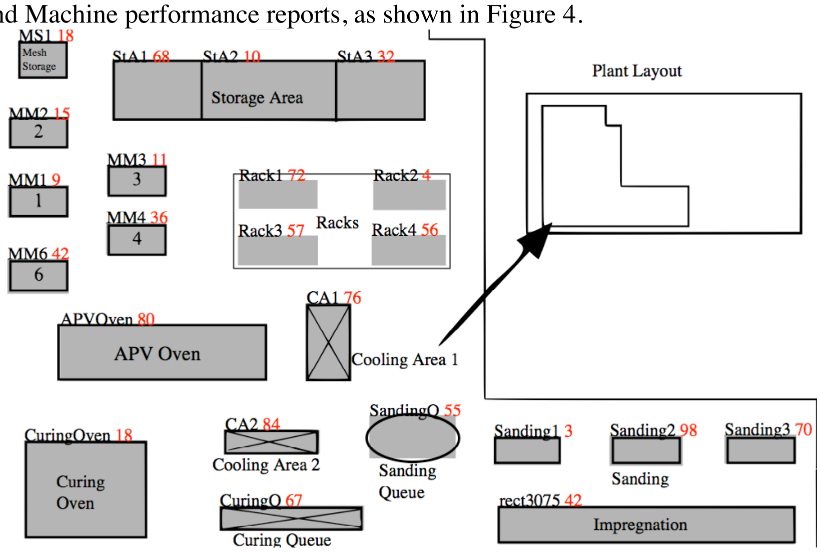

Figure 4 -Floor plan visualisation of knowledge instances

Users can also choose to look at the information from the product point-of-view, by clicking on sensitive areas of the engine, which loads a detailed view of the area of interest, enriched with instances from the documents returned as shown in figure 5. The documents are now grouped into different sections, which are shown as shaded areas- the numbers beside each section indicate the number of retrieved documents related to that section.

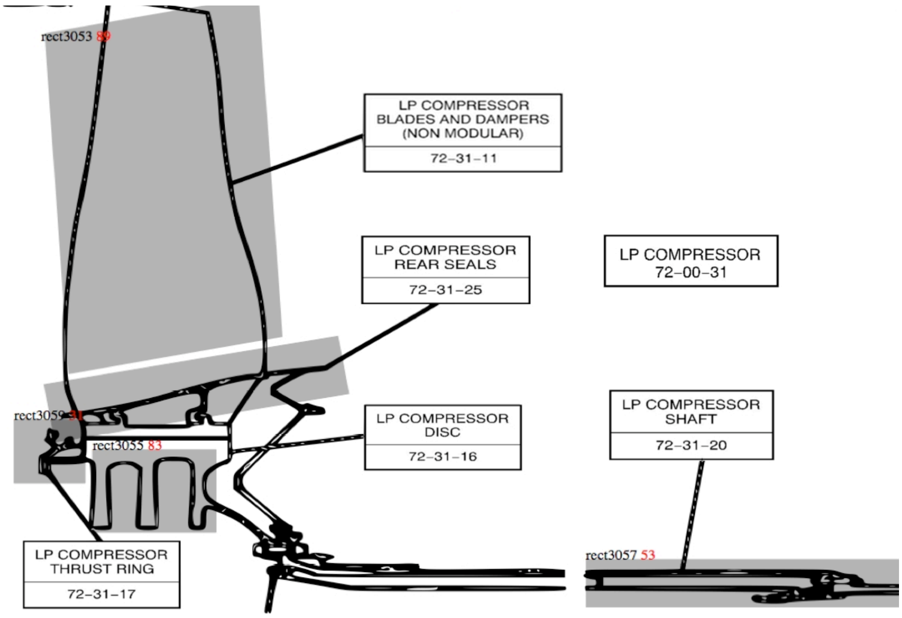

Figure 5 –Detailed view of engine, enriched with knowledge instances

4. Implementation

The system implementation is segmented into a knowledge acquisition and a knowledge exploration system. The knowledge acquisition system is an off-line process implemented in Java. The knowledge visualisation is a web-based dynamic and real-time application, consisting of a javascript frontend and a php backend that communicate using SPARQL queries over a semantic triplestore. The frontend is in charge of interpreting the user interactions and transforming them into corresponding SPARQL queries. For example, clicking on a section of a pie chart would be interpreted as a SPARQL SELECT query. These queries are then transmitted to the backend, which forwards the queries to triplestores. The results from the triplestores are then received by the backend and converted to JSON objects for visualisation in the interface. The system architecture is described in the Figure 6. The block in the right side of the figure shows the front end, while the left side shows the backend processes.

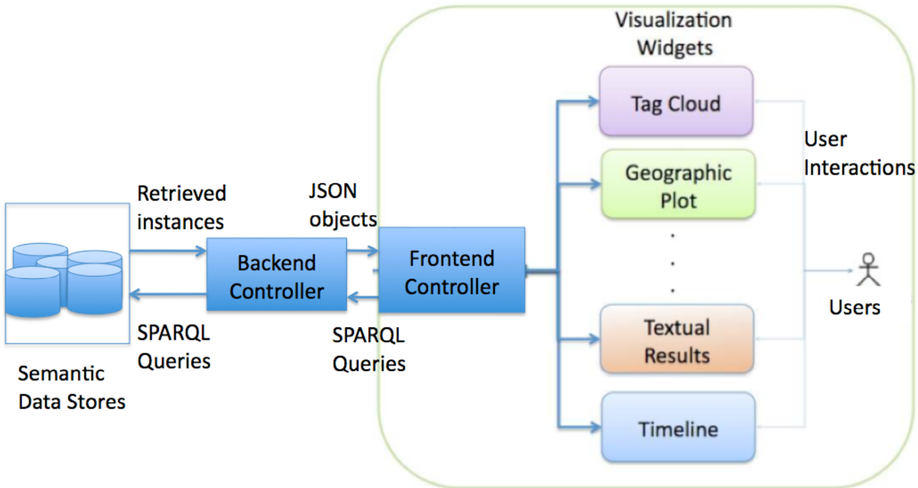

Figure 6 –Visualisation System Architecture

5. Discussion and Conclusions

This paper presented the approach developed during ongoing research work for a project about knowledge management in the manufacturing industry, focusing on how Semantic Web and Information Extraction technologies can be adopted to acquire knowledge from heterogeneous and disparate data whilst providing visualisations to explore, contextualise and aggregate the data, offering multiple perspectives on the knowledge space.

The developed approach is high level and domain independent as it is based on ontologies to structure and visualise knowledge it can be easily applied to a wider context than the manufacturing one. For example it could be applied to any business unit inside a large organisation (i.e. design, service and manufacturing). Expanding the domain will enable organisations to create a large integrated knowledge space available for sharing and reuse.

Future work will concentrate on extending our methodology to different corpora and in enriching the visualisation techniques to better match the user needs. As the project adopts a participatory design paradigm, real users are constantly providing feedbacks on mock-ups and vision demonstrators, to make sure the final prototype will be meeting their needs. This will be complemented by a comparative study of the developed prototype and the current software search systems being used by engineers. Moreover a final user evaluation will be carried out in a real-life scenario to assert the user satisfaction and acceptance of the new technology and a separate in-vitro evaluation will be conducted to test the efficiency and efficacy of the Adaptable IE framework in terms of precision, recall and F-Measure.

Acknowledgments. This research has been supported by the SAMULET project, co-funded by Technology Strategy Board and Rolls-Royce plc. For privacy and security reasons the dataset under analysis is private and all images and data samples have been anonymised.

References

1. Wenger, E. Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, 1998.
2. Bhagdev, R., Chakravarthy, A., Chapman, S., Ciravegna, F., Lanfranchi, V. Creating and Using Organisational Semantic Webs in Large Networked Organisations, In 7th International Semantic Web Conference, Karlsruhe, 2008.
3. Harding, J. A., Shahbaz, M., Srinivas, Kusiak, A. Datamining in manufacturing: A review. American Society of Mechanical Engineers (ASME). Journal of Manufacturing Science and Engineering, 128(4), 969–976, 2006.
4. Wang K. Applying data mining to manufacturing: the nature and implications. Journal of Intelligent Manufacturing of Intelligent Manufacturing. 2007.
5. Choudhary, A., Harding, J., Tiwari, M. Data mining in manufacturing: a review based on the kind of knowledge. Journal of Intelligent Manufacturing, 20(5):501–521–521, 2009.
6. Guh, R. S. Real time pattern recognition in statistical process control: A hybrid neural network/decision tree-based approach. Proceedings of the Institution of Mechanical Engineers. Journal of Engineering Manufacture, 2005.
7. Kwak, C., Yih, Y. Data mining approach to production control in the computer integrated testing cell. IEEE Transactions on Robotics and Automation, 2004.
8. Crespo, F., Webere, R. A methodology for dynamic datamining based on fuzzy clustering. Fuzzy Sets and Systems, 2005.
9. Cunha, D., Agard, B., Kusiak, A. Data mining for improvement of product quality. International Journal of Production Research, 2006.
10. Shahbaz, M., Srinivas, Harding, J. A., Turner, M. Product design and manufacturing process improvement using association rules. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2006.
11. Pan, S. J., Yang, Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359, 2010.
12. Jing, J.. A literature survey on domain adaptation of statistical classifiers. Technical report, 2008.
13. Ciravegna, F., Dingli, A., Petrelli, D., and Wilks, Y. Timely and nonintrusive active document annotation via adaptive information extraction. In Proc. Workshop Semantic Authoring Annotation and Knowledge Management, 2002.
14. Rohrer, R. Visualization and its importance in manufacturing simulation, Industrial Management. 1996.
15. Rohrer, M. W.: Seeing is believing: the importance of visualization in manufacturing simulation. Proceedings of the 2000 Winter Simulation Conference, 2000.
16. Kamath, R. S, Kamat, R. K. Development of cost effective 3D stereo visualization software suite for manufacturing industries. Indian Journal of Science and Technology, 2010.
17. Agrusa, R.; Mazza, V.G.; Penso, R. Advanced 3D Visualization for Manufacturing and Facility Controls., Human System Interactions, 2009.

18. Edgar, G. W. Visualization for non-linear engineering FEM analysis in manufacturing., Proceedings of the 1st conference on Visualization '90, 1990.
19. Gausemeier, J., Ebbesmeyer, P., Grafe, M.,Bohuszewicz, O. v. Cyberbikes - Interactive Visualization of Manufacturing Processes in a Virtual Environment. Proceedings of the Tenth International IFIP WG5.2/WG5.3 Conference on Globalization of Manufacturing in the Digital Communications Era of the 21st Century: Innovation, Agility, and the Virtual Enterprise, 1999.
20. Greif, M. The visual factory: building participation through shared information, 1989.
21. Zhong, Y., Shirinzadeh, B. Virtual factory for manufacturing process visualization, Complexity International, 2008.
22. Stowasser, S. Hybrid Visualization of Manufacturing Management Information for the Shop Floor, Human-computer interaction: theory and practice (Part 2), Volume 2, 2008.
23. Few, S. Information Dashboard Design: The Effective Visual Communication of Data, number 3900693099, O'Reilly Media, 2006.
24. Shneiderman, B. The eyes have it: A task by data type taxonomy of information visualization. In: Bederson, B., Shneiderman, B. (eds.) The craft of information visualization. Morgan Kaufman, San Francisco, 2003.
25. Thorsten Joachims. Estimating the generalization performance of a SVM efficiently. In Proceedings of International Conference on Machine Learning, 2000.
26. Butters, J. and Ciravegna, F.: Authoring Technical Documents for Effective Retrieval. In Proceeding of Knowledge Engineering and Knowledge Management by the Masses (EKAW), 2010.
27. Ackoff, R. L. From Data to Wisdom. Journal of Applied Systems Analysis, 16, 1989.
28. Blei, D., Ng, A. and Jordan, M. Latent Dirichlet Allocation. Journal of Machine Learning Research, 2003.
29. Honglei Guo, Huijia Zhu, Zhili Guo, Xiaoxun Zhang, Xian Wu, and Zhong Su. Domain adaptation with latent semantic association for named entity recognition. In Proc. HTL-NAACL, pages 281–289, June 2009.