

# An Ontology-based Intelligent Data Query System in Manufacturing Networks

Saeidlou, Salman; Saadat, Mozafar; Jules, Guiovanni

DOI:

[10.1080/21693277.2017.1374887](https://doi.org/10.1080/21693277.2017.1374887)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Saeidlou, S, Saadat, M & Jules, G 2017, 'An Ontology-based Intelligent Data Query System in Manufacturing Networks', *Production & Manufacturing Research*, vol. 5, no. 1, pp. 250-267 .  
<https://doi.org/10.1080/21693277.2017.1374887>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# An Ontology-based Intelligent Data Query System in Manufacturing Networks

Salman Saeidlou <sup>a</sup>, Mozafar Saadat <sup>a</sup>, Ebrahim Amini Sharifi <sup>b</sup>, Guiovanni D. Jules <sup>a</sup>

*<sup>a</sup> Department of Mechanical Engineering , School of Engineering, University of Birmingham, Birmingham, United Kingdom; <sup>b</sup> School of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran*

Corresponding author: Salman Saeidlou Email: [sxs989@bham.ac.uk](mailto:sxs989@bham.ac.uk)

Salman Saeidlou, School of Mechanical Engineering, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT, [sxs989@bham.ac.uk](mailto:sxs989@bham.ac.uk)

Mozafar Saadat, School of Mechanical Engineering, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT, [m.saadat@bham.ac.uk](mailto:m.saadat@bham.ac.uk)

Ebrahim Amini Sharifi, School of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran, 158754413, [Amini.sharifi@aut.ac.ir](mailto:Amini.sharifi@aut.ac.ir)

Guiovanni Desire Jules, School of Mechanical Engineering, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT, [gdj039@bham.ac.uk](mailto:gdj039@bham.ac.uk)

# **An Ontology-based Intelligent Data Query System in Manufacturing Networks**

This paper investigates the development of an intelligent data query framework through the use of semantic web technologies for manufacturing purposes. The primary objectives of the ontology-based data query were to develop an efficient and scalable data interoperability and retrieval system; in order to find the most relevant query results with minimum message cost, most hits per query and least response time. This document explains the idea of ontology and the application of the same in the manufacturing domain. A computer simulation software was developed based on a real case study of distributed networks of manufacturing workshops. In this research, a semantic query algorithm was developed where query results are returned by investigating the semantic richness of each workshop. Results were compared with a semantic-free search mechanism based on key performance indicators. The results show the validity of the proposed model for efficient data query when compared to random search.

Keywords: ontology; manufacturing data query; data interoperability; distributed systems

## **1. Introduction**

Evolution of the web from the prevailing web technologies and web services to the semantic web technologies and semantic web services is rapidly advancing. This is as a consequence of the huge amount of data available, the increasing concern of missing out on using valuable data and the cost of using wrong data for decision making (Bikakis, & Sellis, 2016; Esposito *et al.*, 2015). The distinction between data and information is important. Data can be seen as raw, unstructured facts that seem random and useless until organised and information is organised data presented in a given context (Chen *et al.*, 2014). Although missing data (non-response) can be ignored resulting uncertainty in any system, it should also be considered as an Informative

output. It can be argued that a failure by some entities of the system to respond to some queries can also be considered as a response as it provides us with some information. Therefore, in such an example, an absence of data provides information (informative non-response), and this is important in instances where there is minimal data. Therefore, data has potential for a variety of valuable but also erroneous informative outputs, while information is limited to one informative output. Manufacturing currently uses a mixture of data and information infrastructures such as databases of raw data collected from its internal activities and a web of information for business-to-business information transfer along the supply chain.

The impact of the quality of an organization's data on their revenue is addressed by various research studies. According to an EBiz article, there is a dramatic 40% growth in business data each year, while approximately 20% of that data is irrelevant and useless (Tibbetts, 2012). On the other hand, 70% more revenue can be generated by an organization purely based on the quality of its data (SiriusDecisions, 2015). To quantify the cost of data quality, the sales and marketing research firm SiriusDecisions proposes the 1-10-100 rule as a rule of thumb. The rule states that the verification of data as it is entered would cost \$1, data cleansing and removal of duplicates would cost \$10 and the cost of doing nothing about the data would be \$100 (SiriusDecisions, 2015). The development in semantic web technologies such as Resource Description Framework (RDF) instead of traditional Hypertext Markup Language (HTML) allows data to be linked instead of linking information. There is a trend in intelligent systems for embedding RDF descriptions inside HTML pages through RDF attributes (RDFa). The embedded RDF model inside an HTML web page enables automated knowledge discovery and inferencing for intelligent agents (machines). The relationship between data is described by the RDF and the description is stored within a knowledge-base

instead of a database. This knowledge-base has an ontology at its foundation. When data is stored with respect to an ontology, it impacts positively on data re-usability and maintenance.

Furthermore, the ontology enables semantic web services to be developed. Given a simple example of an online query, the traditional web services will support and search only for semantic-free keywords. A semantic web service version of this would be a unified and inferred list of related semantics. Likewise, in manufacturing context, the semantic web services can increase the visibility of good data in an automated way. The semantic web and its services are scalable to accommodate virtually any knowledge in the manufacturing environment, be it predictive production scheduling, production control or design specifications.

On the other hand, there is a trend for manufacturing companies to shift from product-oriented production to user-oriented production (Pine, 1993; Tseng *et al.*, 1996; Tseng & Hu, 2014). The shortening of products' life cycles, global competitiveness and the economy are contributing to the market volatility (Kernschmidt *et al.*, 2015). The shortening of a products' lifecycle is caused by the ability of the competitors to quickly react to release innovative products. The global competition among companies competing on the performance dimensions of product features, price, lead time and quality and also competing within the same market segment are becoming increasingly fierce. To stay on the leading edge, manufacturers need to address the market in new ways (Ageron *et al.*, 2012).

Effective and efficient knowledge sharing and reuse within collaborative manufacturing enterprises is a game-changing opportunity for manufacturers to stay competitive in such a dynamic environment. A seamless flow of information enables manufacturers to increase their throughput while keeping overhead costs to a minimum

(Lin *et al.*, 2011). The manufacturing systems play a crucial role in supporting the manufacturer's efforts as they are intended to provide accurate data in a fast way (Bi, Da Xu, & Wang, 2014). This data comes in a variety of formats, including documents, graphics, video files and the non-human interpretable signals produced by sensors. Data is collated within distributed databases, which are in turn kept within a variety of storage formats and programmes (Song *et al.*, 2013). The data may be retrieved using limited key search attributes within the database, such as Relational Data Bases (RDB) table, row, column, primary and foreign key (Astrova, 2009; Cerbah, 2010; Sane, & Shirke, 2009) . The power of the database's search capacity may increase exponentially if the data is linked, allowing for further attributes and for queries to be combined.

In regard to shop floor operations, the collected data may provide key metrics to the management systems on the shop floors. Such key metrics include level of asset utilisation, statistical control processes and job tracking among others. The underlying intelligence in such systems utilises all available data and provides analysis, forecasting future states and support in decision making (Khan *et al.*, 2011). This ensures that manufacturers make better decisions at all times in supporting their operations (Lee *et al.*, 2014). The correct application of this intelligence gives leverage to allow manufacturers to query their manufacturing sites, to make informed decisions and in turn, perform necessary activities in an educated manner. These activities may include design experimentation, alerts within the system, optimisation, simulation, modelling, prediction and reporting more generally (Davenport *et al.*, 2010).

This project is motivated by the lack of a systematic and intelligent data query system within the manufacturing supply chain. Better decision making resources with intrinsic manufacturing intelligence emerges from a synchronised data sharing system for the mutual benefit of every stakeholder. In this paper, a novel ontology-based query

methodology is presented where the result of each query from the system is evaluated based on the semantic relevance and distance of the query to the local manufacturing documents. The model is built upon a real case study and the results are compared with a semantic-free random walk mechanism. By investigating an intelligent and efficient data query system in a manufacturing network, our work was able to fill a gap in current research.

## **2. Research Methodology**

Intelligent data query is an emerging field of research for the manufacturing supply chain community, as the need for an efficient and effective data sharing and retrieval system is more demanding than ever. This section addresses the key concepts used and the foundation for an intelligent semantic query system.

### **2.1 Manufacturing Ontology**

Ontology is defined as a data model which provides information about a domain with the aid of relationships established among various concepts under that particular domain (Gruber, 1993; Guarino *et al.*, 2009). Organizational knowledge is encapsulated to build a set of knowledge-bases that can combine details such as those of rules, policies, processes, definitions and relationships. Ontology has gained application in many different areas and domains, including manufacturing. Simple to complicated processes are easily explained by ontology in the manufacturing domain; this makes it simpler to create patterns, algorithms and frameworks for the automation of any process.

The Process Specification Language was proposed by Schlenoff *et al.* (2000) from the National Institute of Standards and Technology (NIST) to define a taxonomy to capture and exchange discrete manufacturing process information. Process

Specification Language (PSL) is a neutral standard language for process specifications; such as production scheduling, manufacturing process planning, workflow, business process reengineering, simulation, process modelling and project management. The neutrality of PSL allows process specifications to be translated to and from software-proprietary file formats.

For the domain problem of keeping track of the cost in the manufacturing process, Lemaignan *et al.* (2006) proposed an upper ontology in the manufacturing domain to be a common ancestor for more domain-specific ontologies. MASON (Manufacturing's Semantics Ontology) addresses a similar problem domain to that from the NIST. The difference is that PSL is written in Resource Description Language (RDF); while MASON uses the more up-to-date semantic language called Ontology Web Language (OWL). Moreover, MASON shows a concrete application in expert systems for automatic cost estimation and in semantic-aware multi agent systems. Our work on the other hand looked at an intelligent and efficient data query system in a manufacturing network which is missing in the literature.

In an extensive and rigorous review of the literature, Sanfilippo and Borgo (2016) have focused on the semantics of feature notions. They have highlighted that the current feature-based CAx systems such as Computer-Aided Design (CAD), Computer-Aided Engineering (CAE), Computer-Aided Manufacturing (CAM), and Computer-Aided Process Planning (CAPP) are missing a common understanding of what counts as a feature and the way features are conceptualised. They have emphasised the product knowledge conceptualisation problems within current product life cycle management (PLM) systems and the shortcomings of existing unifying approaches. The proposed ontology-based product knowledge representation would allow reliable data integration while preserving the intended semantics. This will give an insight into the importance of



semantic web technologies for the current manufacturing systems. In our paper, we have tried to demonstrate the advantages of using ontologies for data query in a quantifiable manner.

In another study, Wei *et al.* (2009) proposed an ontology-based approach to address manufacturing design problems. Their system is capable of integrating the whole design process within the product knowledge management to support locating, reusing, integrating and transferring the design knowledge. Their work can be improved and extended by developing more diverse domain-specific engineering ontologies to get a more extensive multidisciplinary integration. In contrast, the presented data query system in our paper is not limited to one aspect of manufacturing supply chain such as product design. The proposed system, designed in a scalable manner to accommodate data integration to the ontology within different tiers of manufacturing domain.

Lin and Harding (2007) have proposed a Manufacturing System Engineering (MSE) ontology which acts as a mediated meta-model for inter-enterprise collaboration. Their ontology is structured into classes, properties and instances. However, this method still involves time and cost overheads, as the individual partners must map their vocabularies to the MSE ontology at the outset; this can be a slow process with the current manual methods. Therefore, this research is limited due to this ineffective process, which may be a major obstacle for its large scale use in information integration in a global supply chain. To ease this problem, formal mapping could become semi-automated via algorithms and heuristics identifying the shared characteristics between the two ontologies.

Usman *et al.* (2013) presented a practical Manufacturing Reference Ontology (MRO) for product design and production domain. Their work is motivated by the lack of a core set of manufacturing concepts for unification of terminologies across various

strands of manufacturing domain. The emphasis was on the formalisation of these core concepts for a semantic-aware system where the MRO will support the development of domain-specific manufacturing ontologies. Their proposed ontology was tested by a qualitative method of inserting two types of facts in the system, namely: facts which are violating the formal classifications of the concepts; and those which conform to the formal classification. The experimentation part was mainly focused on the *Feature* concepts as with the work carried out by Sanfilippo and Borgo (2016). In this study, we looked at a semantic-aware query system in a manufacturing network rather than a comprehensive ontology development which we believe is missing within the manufacturing research community. The contrast between the work of Usman *et al.* (2013) and Lin and Harding (2007) and our work is that our paper made use of the ontology editor Protégé to create an ontology as a tool for data query purposes, while the scope of their work is mainly focused on the development of an efficient and practical core manufacturing ontology.

In a separate research study, Cai *et al.* (2011) presented an ontology-based system called ManuHub which acts as a mediator for manufacturing service providers by efficient and automatic retrieval of the required manufacturing services. Their work is based on Service Oriented Architecture (SOA), and there are two limitations in their work. Firstly, the question as to how to merge their ontology into existing upper-level and domain-specific ontologies in the real world, and secondly how to test their idea on a dynamically changing manufacturing data set.

In order to develop a simulation platform for the formation process of networks of manufacturers, Jules *et al.* (2013, 2015) have developed an ontology based on the Product-Resource-Order-Staff Architecture (PROSA), to model the inter-enterprise

communication during the formation process. The ontology also models the process of resource auctioning which is part of the formation process.

Saeidlou *et al.* (2016) have proposed a hybrid search algorithm where learning process is combined with data extraction from the ontology. This learning technique leads to a more intelligent data query system. The framework used set the foundation for the following research where an intelligent ontology-based data query mechanism is compared with a semantic-free one as well as the concept proposed by Himali *et al.* (2012).

For the scope of this paper, a manufacturing ontology is built upon Protégé ontology editor software, which is a free and open-source leading ontological engineering tool. It provides interface with other knowledge-based tools such as Java Expert System Shell (JESS) and is compatible with various ontology languages and formats such as eXtensible Markup Language (XML), DARPA Agent Markup Language (DAML) and Ontology Inference Layer (OIL) (Ga, 2009). Another reason why this editor is suitable for this research is that it has deductive classifiers for validation of the ontology consistency. Furthermore, it can also export into other formats such as RDF, which is the basis of this research. When modelling in a domain, developers must be able to focus more on the concepts and relations rather than the syntax of the final results. This can be achieved through a Protégé-based editor, resulting modelling at a conceptual level (Noy, 2001; Gennari, 2003).

The concepts of the ontology and the instances are created based on the data from the case study. The scope of the ontology was limited to the workshops' attributes, such as their core competency operations. Some of the key extracted attributes from the case study are summarised and tabulated in section 2.3, which is the foundation for the developed ontology. However, our literature research highlighted that the existing

ontologies do not cover the terminologies used in this data query system research as a result of their ‘domain specific application’ nature. For instance, the concept of ‘quality control’ is not demonstrated in the open source MASON ontology as it was specifically designed for the automatic production cost estimation. In order to widen the scope of the project, the ontology will be further developed for future work and enriched by mapping its main concepts to famous WordNet (<https://wordnet.princeton.edu>) lexical ontology and MASON in an offline manner. This will ensure that the implementation process is facilitated at scale. The alignments of the ontologies will lead to an improved knowledge-base, which will allow more data to be queried from the system. As an example, for this case study, the concept of ‘production cost’ is not considered as it can be covered by mapping the proposed ontology to MASON. Figure 1 shows the core concepts of the proposed ontology where each core concept is further formalised into more sub-concepts.

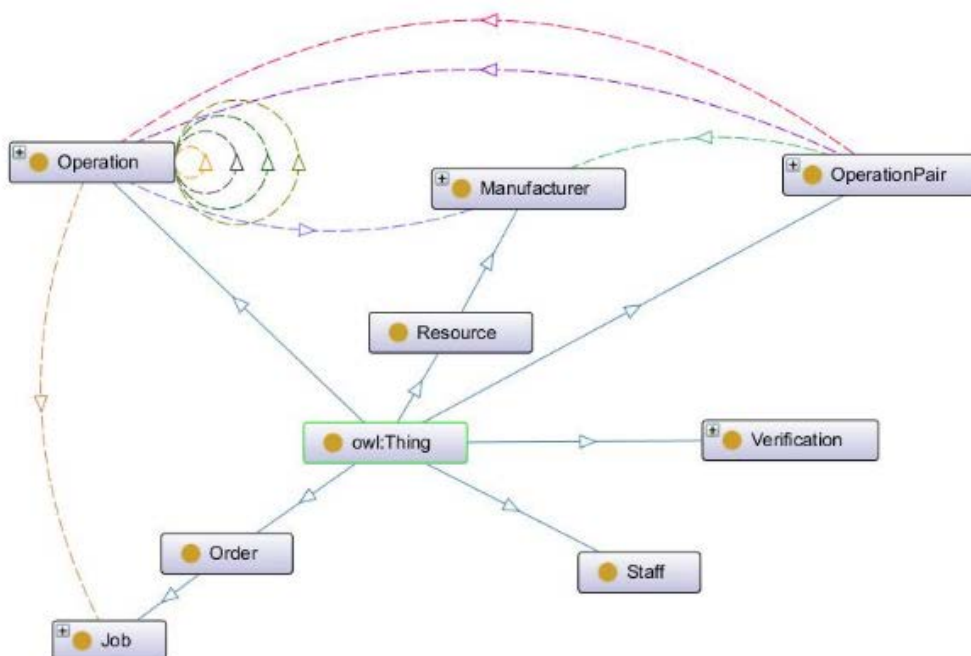


Figure 1. Developed ontology for manufacturing data query

For instance, the Manufacturer concept is a sub class of Resource concept. Similarly, the Manufacturer concept is further specialised into different manufacturer instances, where each one provides a set of manufacturing operations. Each Order consists of a set of Jobs which are defined by unique process plans. These process plans outline which operations are required for each job. Therefore, *isNeededBy* define which operations are needed for each job and *isProvidedBy* define which manufacturer provides that particular operation. Figure 2 shows an example of detailed object properties within the developed ontology.

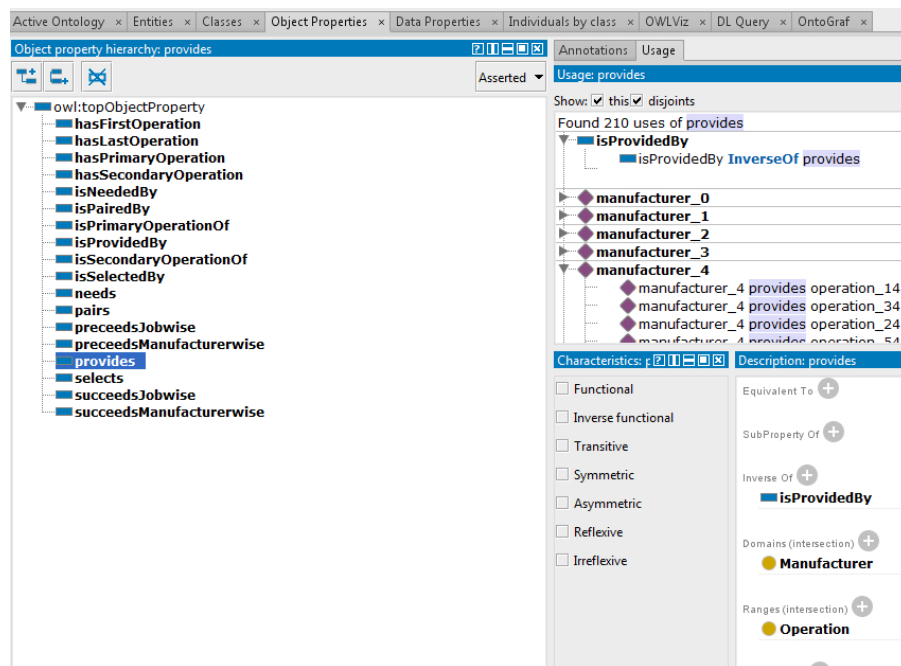


Figure 2. Object properties of the proposed ontology

## 2.2 Industrial Case Study

The case study was carried out at the Gruppo Fabricazione Meccanica (GFM) S.r.l. Group, Italy, and its objective was to use the result to help the understanding of how the theories of semantic data query apply to real manufacturing business situations. The

case study will provide information to create a model of how the company, under study, is involved in the life cycle of its products. To do this, the study will look at the activities involved when a call for a proposal is received; an offer is generated; a firm order is received; material sourcing, production, quality control and transport logistics are triggered; and finally, after sales service is provided. During all these phases, hundreds of queries and messages are generated and as a considerable portion of these are unnecessary and can be avoided with an intelligent query system, they are highlighted.

GFM acts as a central company to a pool of manufacturing workshops and suppliers. They collaborate with over 300 suppliers and over 40 workshops to fulfil customer production needs. Their expertise is in the production of gas turbines, steam turbines and generator parts. GFM is a single point of contact for this study, and the case study data is extracted by queries in Microsoft Structured Query Language (SQL) Server to GFM's management and information system. They have internally developed the software that monitors the supplier's performances so it can better meet their needs, and have built procedures / queries for extracting the data.

### **2.3 *Simulation Software***

In collaboration with GFM, the most important data about each individual supplier is gathered and imported into different nodes in the simulation software namely, the operations provided, workpiece materials, fleet of machines and quantity, types of testing, tolerance dimensions and certifications. These are parsed into JAVA objects and stored in each node as its local database. A JAVA simulation platform is developed in NetBeans integrated development environment (IDE) to implement the model of network used by the manufacturing workshops and to simulate the message generation

process during a data query. The simulation of the manufacturing network is built upon a primitive version of the software presented in a previous work (Jules *et al.* 2015). However, the current simulation is mainly focused on the flow of data within the network and the use of ontology for data query rather than the network formation and scheduling. Each node or object represents one supplier/workshop and has the capability of communicating with other nodes. During the simulation, a query list is generated from one node (GFM). This is then circulated among the nodes until the result of the query is found. The platform makes use of two protocols for the search mechanism: namely random search protocol and ontology search protocol. The random search setting involves randomly selected nodes, while in ontology search protocol the query is sent to nodes with high probability of containing the answer to the query based on the unified ontology. These two search methods are explained in more detail in section 2.4 and 2.5. The schematic of this distributed manufacturing network is illustrated in Figure 3.

Table 1. Examples of key extracted attributes of workshops

WORKSHOP'S ATTRIBUTES	SAMPLE VALUES
SUPPLIER CODE / GFM CODE	2.3.4.10.370 / 2AE001031.00
SUPPLIER TYPE	MACHINING ; TREATMENT ; RAW MATERIAL PROVIDER
OPERATION OFFERED	TURNING ; MILLING ; HEAT TREATMENT
FLEET OF MACHINES	MAZAK NEXUS 250MY (MAX TURNING DIAMETER: 70 MM)
CERTIFICATIONS	ISO 9001:2008; CERTIFICATE NO. 6391; EXPIRY DATE: 28/04/2018
NON-DESTRUCTIVE TESTING (NDT) CONTROL	PENETRATION TEST; HARDNESS TEST; VISUAL TEST
NON-DESTRUCTIVE TEST	RADIOGRAPHIC INSPECTIONS; DYE PENETRATE INSPECTIONS; MAGNETIC PARTICLES INSPECTIONS
DESTRUCTIVE TEST	TENSILE TESTING; IMPACT TESTING; SPECTROGRAPHIC EXAMINATION; STATIONARY SPECTROMETER
MATERIALS	21CRMOV57; X15CR13; 26NICRMO146
STABILISATION	WTLV9520-01; WTLV9520.01
SOLUBILISATION	WTLV9520-01; 424W834R0; WTLV9520.01
NUMBER OF PLACED ORDERS PER MONTH/ PER YEAR	JULY 2006 / 6 ORDERS
REQUESTED DELIVERY DATE	10/10/2006
ACTUAL DELIVERY DATE	04/12/2006
DELAYS	55 DAYS
QUALITY	CONFORM; NON-CONFORM; NUMBER OF LOTS CONFORM

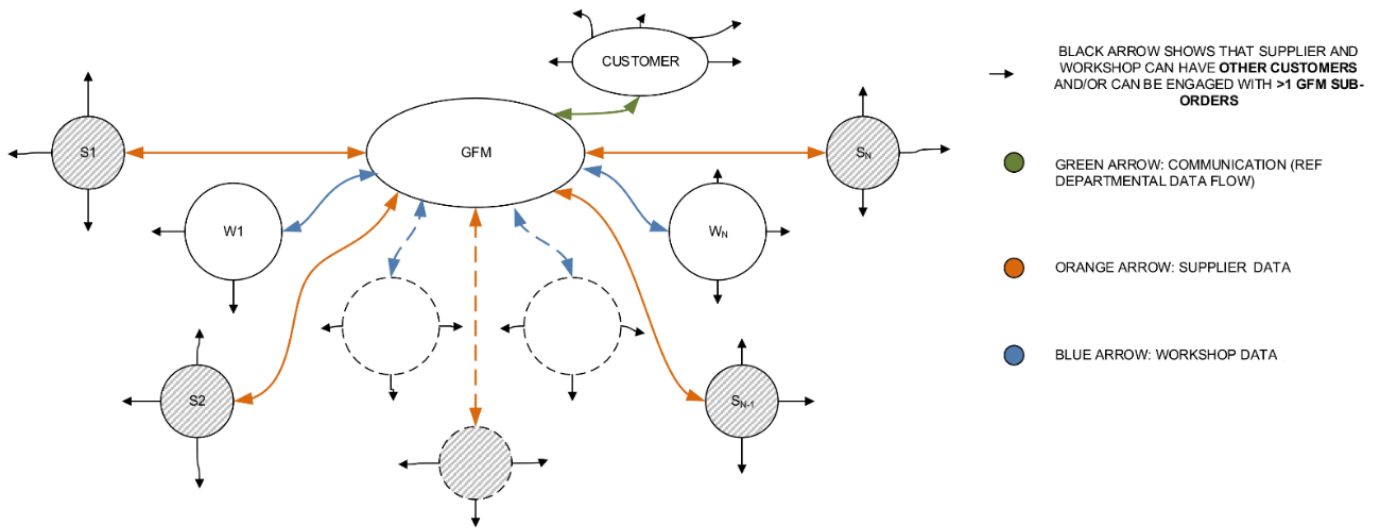


Figure 3. Case study network structure (Data flow between network entities)

This figure illustrates the formation of a network and the data which support the network operation for 1 sub-order. This network is constituted by supplier and workshop data, which store data at every instance. The green, orange and blue arrows present the customer communication, supplier and workshop data flow within the network respectively. As the queried data cannot be accessed easily as it can be stored in any section of the network, ontology is the most efficient solution to access any data within the network.

#### 2.4 Random Search Protocol

In a manufacturing network, each workshop has a myopic view of the system and the data stored in different resources. Search protocols enable finding the best queried resources by routing the query to the right workshop. This research is trying to show the potential of ontology for the development of an intelligent and efficient query system



compared to a semantic-free search method. The random search method adapted in this study is based on the Gnutella algorithm (Breadth-First Search (BFS) algorithm), which uses flooding to search all the workshops for a given query (Yeferny *et al.*, 2011; Kapoor *et al.*, 2013; Arour & Yeferny, 2015). This method is well documented, and it is used due to its simplicity of implementation. The aim is not to develop a random search protocol, but to modify this protocol for the application within a manufacturing network and compare it with the proposed novel ontology search protocol.

The search mechanism is built upon two protocols, namely the random search protocol and the ontology protocol (outlined in section 2.5). During programme initialisation, an object called *Message* is created containing some properties such as *QueryList* and *Hit*. *QueryList*, as the name suggests, contains the query that is sent to the network, and *Hit* is a boolean indicating whether or not the query result is found. The simulation software is developed on a cycle-based evaluation mechanism. After each cycle, if the *Message* finds the answer, it triggers *Hit*; otherwise it is sent to a random neighbour node and also adds to the counter *CrossStep*. Moreover, each protocol has properties called *LastMessage*, *IncomingMessages* and *ProcessList* which will be used in the *Observer* method for the evaluation of the whole process. This performance measure evaluation is explained later in the following sections. In each cycle, the content of the *QueryList* is compared with the local storage of each workshop. For each common match between the *QueryList* and the local storage data list, a counter is added inside the *FindRelevance* method and that relevance is compared with a predefined threshold. If the value of the relevance exceeds the threshold a *Hit*; is triggered, otherwise the cycle proceeds. The algorithm for the random search protocol is demonstrated in Algorithm 1 in Figure 4.

<p>Algorithm 1:</p> <ol style="list-style-type: none"> <li>1- QueryList: Create a list from queries and send it to nodes via message</li> <li>2- There is a variable in each node called LastMessage resembling the last received message</li> <li>3- ProcessList: The list of operations offered by each node(workshop)</li> <li>4- Relevance: The value obtained by comparing LastMessage and QueryList</li> <li>5- Hit: The Boolean value indicating receive of message from a node(workshop)</li> <li>6- Threshold: Hit occurs if the value of Relevance is higher that Threshold</li> <li>7- CrossStep: Number of hops that a message travel until it trigger Hit</li> </ol>
<pre> Procedure: If (lastMessage != null) { //Compare LastMessage and QueryList and return Relevance value as the result Relevance = 0.0;   foreach (queryItem in QueryList){     if (ProcessList.contains(queryItem)) {       Relevance ++;     }   } //Hit the message if the value of Relevance is higher than Threshold, otherwise send the message to a random neighbor If(Relevance&gt; Threshold) { Message.hit=true; } Else { Message.hit=false; Message.crossStep++; NeighborsPeer= getNeighbors(Random integer Index); SendMessage(Message, NeighborsPeer) } } </pre>

Figure 4. Random search protocol

## 2.5 *Ontology Search Protocol*

As part of ontology engineering practice, there are different ways to create and represent ontologies. The common key standards are Web Ontology Language, also known as OWL and Resource Description Framework known as RDF. In the ontology search protocol, RDF has been used for the representation of domain information. Abstract syntaxes are used for the description and are linked with the concrete syntaxes of the formal semantics. The documentation part in this case involves XML serialisation as well as the formal semantics. Graphical representation in the form of a data model of the RDFs comprises a subject, a predicate and the target object, which are presented with the help of nodes and a directed arched diagram.

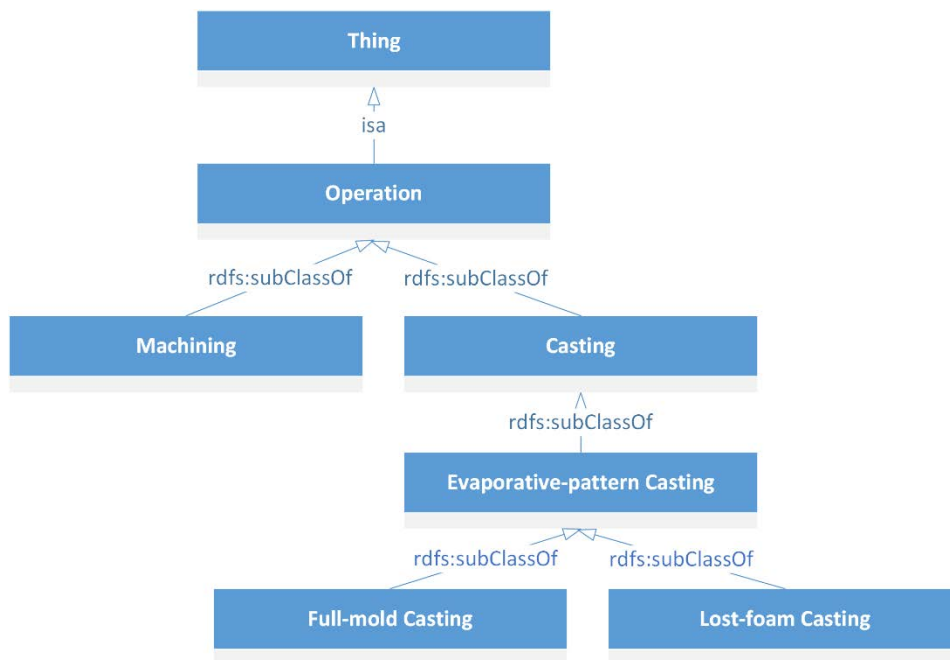


Figure 5. RDF representation of manufacturing ontology

In this research, the semantic distance of domain concepts is calculated from the RDF graphs and was used during the search process for the most relevant results. If the Operation concept from Figure 1 is further expanded to its sub-classes the relation between existing operations can be investigated. For instance, Machining, Casting, Forming and Joining are all sub-classes of Operation. The Casting itself has sub classes such as Full-mold casting and Lost-foam casting. The RDF graph of such a hierarchy can demonstrate the relationship between these concepts in the ontology and also shows the semantic distance of these concepts. As an example, the semantic distance between Lost-foam casting and Full-mold casting is two, as it is shown in Figure 5. Algorithm 2 elaborates on the procedure involved when ontology protocol is in use.

<p>Algorithm 2:</p> <ol style="list-style-type: none"> <li>1- QueryList: Create a list from queries and send it to nodes via message</li> <li>2- There is a variable in each node called LastMessage resembling the last received message</li> <li>3- ProcessList: The list of operations offered by each node(workshop)</li> <li>4- OntologyRelevance: The value obtained by comparing LastMessage and QueryList form RDF graph and indicating the semantic distance of two concepts</li> <li>5- Hit: The Boolean value indicating receive of message from a node(workshop)</li> <li>6- Threshold: Hit occurs if the value of OntologyRelevance is higher that Threshold</li> <li>7- CrossStep: Number of hops that a message travel until it trigger Hit</li> </ol>
<pre> Procedure: If (lastMessage != null) { //Compare semantic distance of two concepts from QueryList and LastMessage and //return the OntologyRelevance value OntologyRelevance = 0.0; foreach(QueryItem in QueryList) {      foreach (ProcessItem in ProcessList) { //The higher the distance number, the farther semantic distance of two concepts OntologyRelevance += 10 - getOntologyDistance(QueryItem, ProcessItem);     } } //Hit the message if the value of OntologyRelevance is higher than Threshold, //otherwise send the message to the closest neighbor in the RDF graph with the value more than Threshold If(OntologyRelevance &gt; Threshold) {     Message.hit=true; }Else {     Message.hit=false;      While(neighbor!=null)     {         if (getOntologyDistance (ListQuery, neighbor.processList) &gt; Threshold) {              Message.crossStep++;             SendMessage(Message, neighbor)         }else {              Neighbor= neighbor.next();         }     } } } </pre>

Figure 6. Ontology search protocol

The contribution of this research could be put into perspective as follows: (i) the proposed ontology search protocol presents a novel data query system, which focuses on retrieving information with respect to the semantic content of a workshop and by considering all the relevant accessible data within a workshop; (ii) this data query algorithm will evaluate the semantic richness of the database to return the most relevant results for each query, in comparison to a semantic-free based approach, even if the answer to the query does not exist within the documents with the help of reasoning; and

(iii) the development of the proposed search mechanism results in a significantly more efficient data retrieval system than the other data query systems such as Random Walk and OSQR (Himali *et al.*, 2012) considering the recall and precision metrics of the search quality from the literature.

## **2.6 Performance Evaluation**

Evaluation of the performance of a data retrieval system is a crucial task which entails a well-established test method and benchmark. The results of the queries from the ontology will be evaluated against the actual existing data from the manufacturing environment in order to determine the fidelity of the ontology query model and the simulation model. The Key Performance Indicators (KPIs) involved in calculating fidelity of the proposed system are devised based on Precision and Recall performance criteria. Based on the literature review, these KPIs are the most prominent performance measures when dealing with information retrieval systems and ontologies (Raghavan *et al.*, 1989; Müller *et al.*, 2001; Euzenat, 2007), and therefore they have been chosen for the evaluation of this research. Furthermore, these KPIs have been chosen to allow the comparative approach to other similar research papers (Himali *et al.*, 2012). In the configuration of the simulation software, an Observer method is created to evaluate the efficiency of the algorithms developed. The Observer calculates the number of Hits from all incoming messages, message cost, maximum CrossSteps and the sum of all CrossSteps, and these are the basis of the performance evaluation. The message cost is the number of all generated messages, and the maximum CrossStep is the longest distance travelled by the Message to trigger the Hit. The Response time is considered as the first occurring Hit. Precision is a measure of returned query relevancy, which is defined as the number of relevant retrieved instances to the total number of instances; and the other measures concern hits per query and recall. These are defined as the ratio

of the number of relevant records retrieved to the total number of relevant records in the system. All these key performance indicators are used in this research to assess the basis of the proposed protocols.

In addition to the above, in every information retrieval, two main metrics of Precision and Recall are targeted to measure the classifier output quality of search. These metrics are also linked to their harmonic mean, F-Score, which can be obtained by the following formulae: (<https://www.crosswise.com/cross-device-learning-center>):

$$Recall = TP / (TP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

Precision is mainly known as the measure of result relevancy, which is dependent on the number of true positives (TP) and false positives (FP). On the other hand, recall is a metric value to evaluate the number of truly relevant results obtained and it is dependent on the number of True positives (TP) and False negatives (FN). The link between these metrics is a great illusion of the nature of the current problem in the process of information retrieval. By recalling data, all relevant documents can be accessed and finally sorted through junk, and this is interpreted as precision. In an information retrieval system, low precision leads to a high number of incorrectly predicted results, and low recall leads to a lower number of correctly predicted results. Therefore, the ideal system would have to demonstrate high precision and recall to achieve its most accurate results. As accuracy is very misleading in the evaluation of the quality of an information retrieval, these metrics are taken into account together to ensure a high quality of information retrieval.

### 3. Results and Discussion

The results of the simulation are depicted in the Figures 7-11 and are based on the aforementioned evaluation criteria. As demonstrated in Figures 7 and 8, the ontology protocol shows a considerable amount of improvement in the number of travelled steps by the Message. The random search mechanism travels more distance among the nodes to find the query result. As illustrated in Figures 7 and 8, there is a maximum difference of more than 200,000 in the comparison of total cross steps, as well as a maximum change of over 1,500 maximum cross steps as a result of the new proposed ontology. Figures 9 and 10 show experimental results for Hit and Hit per Query. As depicted in these Figures, the ontology protocol shows a dramatic improvement of 166.67% and 7,900% in total Hit counts compared to OSQR protocol and Random protocol respectively (Himali *et al.*, 2012). As demonstrated in figures 9 and 10, the ontology protocol has been able to reach a maximum of 7,920 Hits per Cycle, while the random protocol holds only 99 Hits. The recall rate of the proposed search mechanism is drastically higher than both the OSQR and Random protocols respectively. However, the OSQR protocol was based on WordNet ontology which used a wide range of sample data inputs. In order to accurately compare the results of the proposed protocol with OSQR, it is essential to simulate the constraints by equalising the number of data inputs.

Figure 11 shows how the ontology protocol outperforms the random search method and OSQR in regards to precision rate. With the proposed algorithm, the possibility of an incorrect identification of data is minimised to ensure the least mismatching of the relevant data with irrelevant information to the query. The calculated improvement rate of precision resulting from this ontology is estimated at 228.57% compared to OSQR and 604.10% compared to Random search, which points out the efficiency of ontology protocol when considering hit messages per query. Table 2 tabulates the key

performance evaluation criteria.

Table 2. Key performance evaluation criteria

	Recall	Precision	F-Score	Hit Count
<b>Random</b>	0.68	0.98	0.073	99
<b>OSQR</b>	1.49	2.1	0.42	2970
<b>Ontology</b>	12.72	6.9	0.82	7920

The only downfall of the proposed ontology protocol is the response time. For the simulation run illustrated in this paper, the average response time for the ontology protocol is 9 minutes; while the random search is done in under a minute. This is due to the semantic calculation happening in the ontology protocol and could be alleviated in the future work.

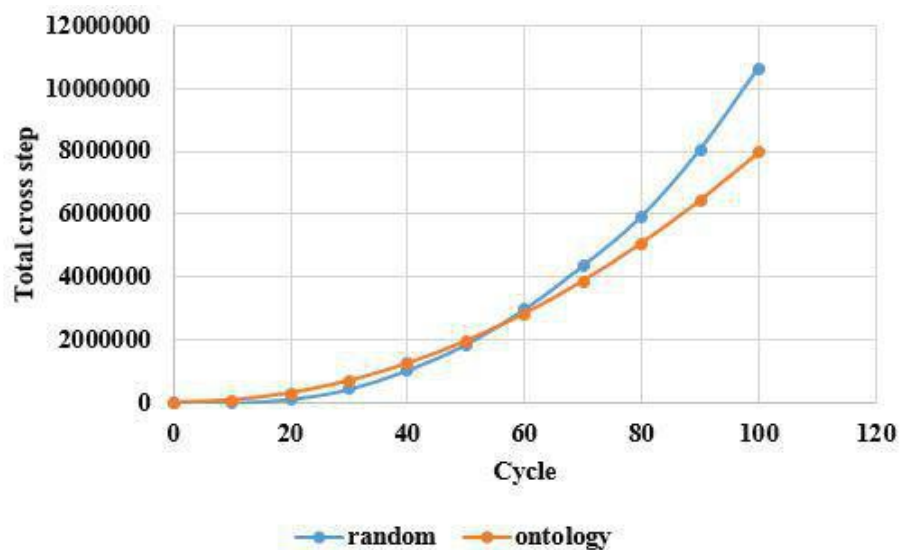


Figure 7. Total cross step comparison of Random search and Ontology



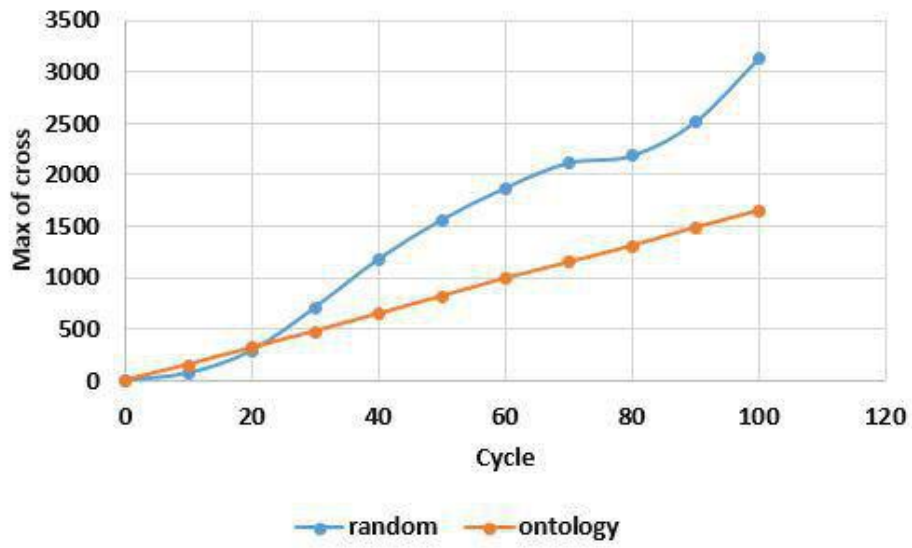


Figure 8. Maximum cross step comparison of Random search and Ontology

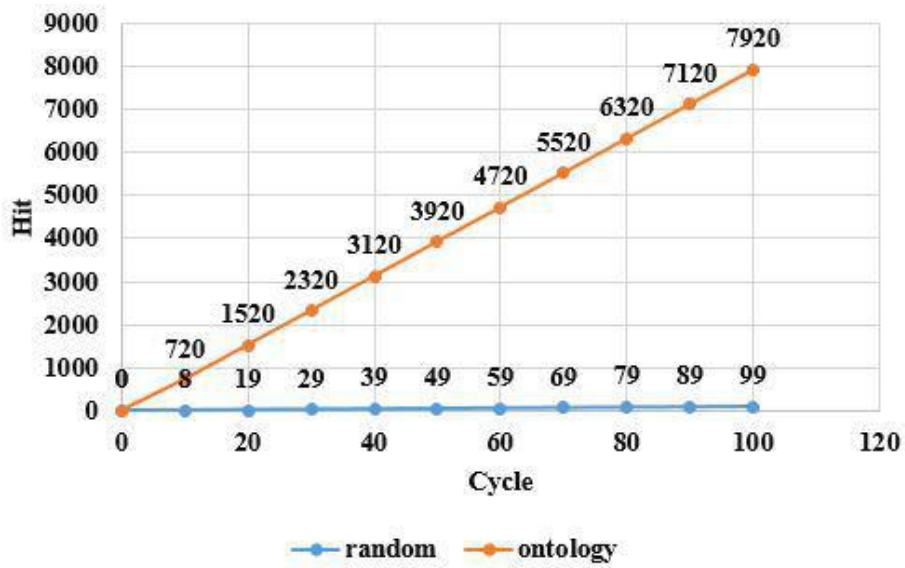


Figure 9. Hit rate comparison of Random search and Ontology

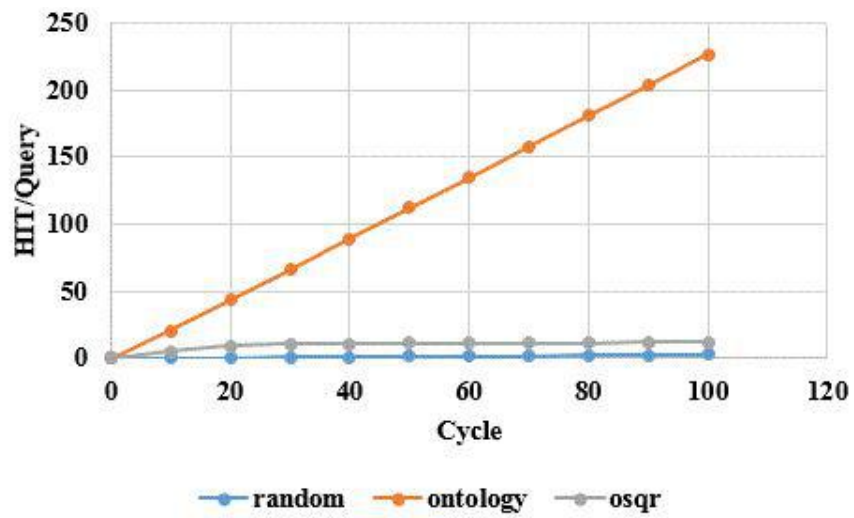


Figure 10. Hit per Query comparison of Random, Ontology and OSQR protocol

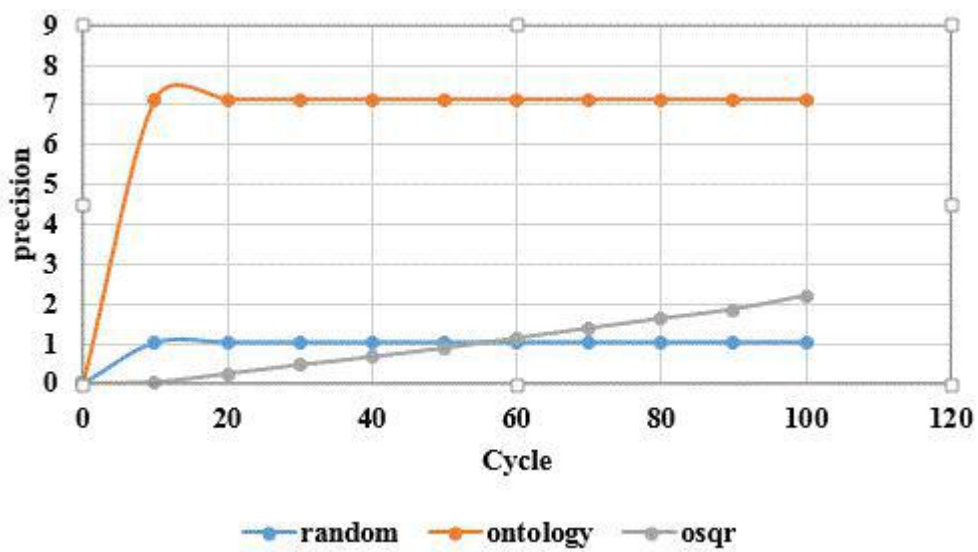


Figure 11. Precision rate comparison of Random, Ontology and OSQR protocol

#### **4. Conclusion and Future Work**

A manufacturing ontology was created in this research to model the knowledge from a real case study. Accordingly, a simulation platform was developed to demonstrate the validity of an intelligent data query system. Based on the results, the effectiveness of the ontology protocol was shown compared to a random search mechanism and the well-established OSQR method from literature. Using the ontology not only helped to reach the desired query results with fewer messages, but it was also useful for capturing the most relevant results per query. Moreover, it will be useful in case of the absence of query results in local storage databases. For instance, if the result for a query does not exist in the nodes, the random search will return no result with a huge amount of generated messages. However, the ontology protocol will return the closest available answer to that query through the use of semantic relations. This will be beneficial especially in a dynamic manufacturing environment.

The scope of this paper was limited to just two search mechanisms and a number of limited evaluation metrics. As future work, the research will focus on the enhancement of the manufacturing ontology by using well-known existing ontologies such as WordNet and MASON. Furthermore, the authors are interested in the development of a multi-agent system where autonomous agents automate the whole query process. They will communicate with each other to find the query results and also to populate the ontology in an automatic way for knowledge reuse.

#### **Acknowledgment**

This research has been supported by GFM S.r.l. (Italy). The authors express their gratitude to GFM for their kind support throughout the project.

## References

- Ageron, B., Gunasekaran, A., & Spalanzani, A. (2012). Sustainable supply management: An empirical study. *International Journal of Production Economics*, 140(1), 168-182.
- Arour, K., & Yeferny, T. (2015). Learning model for efficient query routing in P2P information retrieval systems. *Peer-to-Peer Networking and Applications*, 8(5), 741-757.
- Astrova, I. (2009). Rules for mapping SQL relational databases to OWL ontologies. In *Metadata and Semantics* (pp. 415-424). Springer US.
- Bi, Z., Da Xu, L., & Wang, C. (2014). Internet of things for enterprise systems of modern manufacturing. *IEEE Transactions on Industrial Informatics*, 10(2), 1537-1546.
- Bikakis, N., & Sellis, T. (2016). Exploration and visualization in the web of big linked data: A survey of the state of the art. *arXiv preprint arXiv:1601.08059*.
- Cai, M., Zhang, W. Y., & Zhang, K. (2011). ManuHub: a semantic web system for ontology-based service management in distributed manufacturing environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(3), 574-582.
- Cerbah, F. (2010). Learning ontologies with deep class hierarchies by mining the content of relational databases. In *Advances in knowledge discovery and management* (pp. 271-286). Springer Berlin Heidelberg.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Harvard Business Press.
- Esposito, C., Ficco, M., Palmieri, F., & Castiglione, A. (2015). A knowledge-based platform for Big Data analytics based on publish/subscribe services and stream processing. *Knowledge-Based Systems*, 79, 3-17.
- Euzenat, J. (2007, January). Semantic Precision and Recall for Ontology Alignment Evaluation. In *IJCAI* (Vol. 7, p. 348353).
- Ga, D., Djuric, D., & Deved, V. (2009). *Model driven engineering and ontology development*. Springer Science & Business Media.
- Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., ... & Tu, S. W. (2003). The evolution of Protégé: an environment for

- knowledge-based systems development. *International Journal of Human-computer studies*, 58(1), 89-123.
- Gruber, T. (1993). What is an Ontology. *WWW Site <http://www-ksl.stanford.edu/kst/whatis-an-ontology.html>*
- Guarino, N., Oberle, D., & Staab, S. (2009). What is an Ontology?. In *Handbook on ontologies* (pp. 1-17). Springer Berlin Heidelberg.
- Himali, D. R., Navathe, S. B., & Prasad, S. K. (2012, December). OSQR: A framework for ontology-based semantic query routing in unstructured P2P networks. In *High Performance Computing (HiPC), 2012 19th International Conference on* (pp. 1-10). IEEE.
- Jules, G. D., Saadat, M., & Saeidlou, S. (2013, August). Ontological extension of PROSA for manufacturing network formation. In *International Conference on Industrial Applications of Holonic and Multi-Agent Systems* (pp. 140-151). Springer Berlin Heidelberg.
- Jules, G. D., Saadat, M., & Saeidlou, S. (2015). Holonic Ontology and Interaction Protocol for Manufacturing Network Organization. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(5), 819-830.
- Kapoor, H., Mehta, K., Puri, D., & Saxena, S. (2013). Survey of various search mechanisms in unstructured peer-to-peer networks. *International Journal of Computer Applications*, 68(6).
- Kernschmidt, K., Preißner, S., Raasch, C., & Vogel-Heuser, B. (2015, October). From Selling Products to Providing User Oriented Product-Service Systems—Exploring Service Orientation in the German Machine and Plant Manufacturing Industry. In *IFIP International Conference on Product Lifecycle Management* (pp. 280-290). Springer International Publishing.
- Khan, S., Ganguly, A., & Gupta, A. (2011). Creating Knowledge for Business Decision Making.
- Lee, C. K. H., Choy, K. L., Law, K. M. Y., & Ho, G. T. S. (2014). Application of intelligent data management in resource allocation for effective operation of manufacturing systems. *Journal of Manufacturing Systems*, 33(3), 412-422.
- Lemaignan, S., Siadat, A., Dantan, J. Y., & Semenenko, A. (2006, June). MASON: A proposal for an ontology of manufacturing domain. In *IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications (DIS'06)* (pp. 195-200). IEEE.

- Lin, H. K., & Harding, J. A. (2007). A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration. *Computers in Industry*, 58(5), 428-437.
- Lin, L. F., Zhang, W. Y., Lou, Y. C., Chu, C. Y., & Cai, M. (2011). Developing manufacturing ontologies for knowledge reuse in distributed manufacturing environment. *International Journal of Production Research*, 49(2), 343-359.
- Müller, H., Müller, W., Squire, D. M., Marchand-Maillet, S., & Pun, T. (2001). Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recognition Letters*, 22(5), 593-601.
- Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W., & Musen, M. A. (2001). Creating semantic web contents with protege-2000. *IEEE intelligent systems*, 16(2), 60-71.
- Pine, B. J. (1993). *Mass customization: the new frontier in business competition*. Harvard Business Press.
- Raghavan, V., Bollmann, P., & Jung, G. S. (1989). A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems (TOIS)*, 7(3), 205-229.
- Saeidlou, S., Saadat, M., & Aminisharifi, E. (2016, August). Intelligent Semantic Query for Manufacturing Supply Chain. In *Advances in Manufacturing Technology XXX: Proceedings of the 14th International Conference on Manufacturing Research, Incorporating the 31st National Conference on Manufacturing Research*, September 6–8, 2016, Loughborough University, UK (Vol. 3, p. 419). IOS Press.
- Sane, S. S., & Shirke, A. (2009, January). Generating OWL ontologies from a relational databases for the semantic web. In *Proceedings of the International Conference on Advances in Computing, Communication and Control* (pp. 157-162). ACM.
- Sanfilippo, E. M., & Borgo, S. (2016). What are features? An ontology-based review of the literature. *Computer-Aided Design*, 80, 9-18.
- Schlenoff, C., Denno, P., Ivester, R., Libes, D., & Szykman, S. (2000). An analysis and approach to using existing ontological systems for applications in manufacturing. *AI EDAM*, 14(04), 257-270.
- Sirius Decisions, (2015) - *The Impact of Bad Data on Demand Creation* [Online]. Available: <http://bit.ly/2sxn2Y>

- Song, F., Zacharewicz, G., & Chen, D. (2013). An ontology-driven framework towards building enterprise semantic information layer. *Advanced Engineering Informatics*, 27(1), 38-50.
- Tibbetts, H. (2012). *Integration on the Edge: Data Explosion & Next-Gen Integration*. [Online]. Available: <http://www.ebizq.net/blogs/integrationedge/2012/01/>
- Tseng, M. M., & Hu, S. J. (2014). Mass customization. In *CIRP Encyclopedia of Production Engineering* (pp. 836-843). Springer Berlin Heidelberg.
- Tseng, M. M., Jiao, J., & Merchant, M. E. (1996). Design for mass customization. *CIRP Annals-Manufacturing Technology*, 45(1), 153-156.
- Usman, Z., Young, R. I. M., Chungoora, N., Palmer, C., Case, K., & Harding, J. A. (2013). Towards a formal manufacturing reference ontology. *International Journal of Production Research*, 51(22), 6553-6572.
- Wei, S., Qin-Yi, M., & Tian-Yi, G. (2009). An ontology-based manufacturing design system. *Information Technology Journal*, 8(5), 643-656.
- Yeferny, T., Bouzeghoub, A., & Arour, K. (2011). A query learning routing approach based on semantic clusters. *International Journal of Advanced Information Technology*, 1(6), 1.