Supporting Multi-view User Ontology to Understand Company Value Chains

Landong Zuo¹, Manuel Salvadores¹, SM Hazzaz Imtiaz¹, John Darlington¹, Nicholas Gibbins¹, Nigel R Shadbolt¹, and James Dobree²

¹ Intelligence, Agents, Multimedia (IAM) Group School of Electronics and Computer Science University of Southampton, UK {lz,ms8,hsmi,jd,mng,nrs}@ecs.soton.ac.uk ² Semantric Ltd. 25 Landsdowne Gardens London, SW8 2EQ, UK james@semantric.com

Abstract. The objective of the Market Blended Insight (MBI) project is to develop web based techniques to improve the performance of UK Business to Business (B2B) marketing activities. The analysis of company value chains is a fundamental task within MBI because it is an important model for understanding the market place and the company interactions within it. The project has aggregated rich data profiles of 3.7 million companies that form the active UK business community. The profiles are augmented by Web extractions from heterogeneous sources to provide unparalleled business insight. Advances by the Semantic Web in knowledge representation and logic reasoning allow flexible integration of data from heterogeneous sources, transformation between different representations and reasoning about their meaning. The MBI project has identified that the market insight and analysis interests of different types of users are difficult to maintain using a single domain ontology. Therefore, the project has developed a technique to undertake a plurality of analyses of value chains by deploying a distributed multi-view ontology to capture different user views over the classification of companies and their various relationships.

Keywords: Ontology Engineering, Data Management, Market Modeling, Semantic Web, Value Chain, Network Analysis.

1 Introduction

The MBI project [14] has the clear objective to make a significant performance improvement in UK business to business (B2B) marketing activities by providing unparalleled insight into UK business activity and support for intelligent decision-making processes. The project consortium includes the marketing departments of 5 influential companies each representing a different set of needs within the UK B2B market, 3M, AXA, British Gas Business, National Australia Group Europe (Clydesdale and York shire Bank) and Parcel Force Worldwide.

A major issue for the project, and the market analysis techniques that have been researched, is how domain knowledge can be utilized to support the diverse analysis interests from all project partners. The issue is made complex by the need for rich semantics of the specific business scenarios that need to be modeled while at the same time supporting common forms of market analysis such as a value chain analysis. This paper argues that a single domain ontology is too rigid to resolve the problem and therefore, proposes a layered user viewpoint model which gives extra flexibility in terms of unique user views of information classification and relationship definition. The principle of multiple user view ontologies is generic but is illustrated in this paper with the latest achievement of value-chain analysis application.

The remainder of paper is structured as follows: section 2 reviews the related work, section 3 gives the background knowledge of value-chain analysis, section 4 describes the application architecture and processing work flow, section 5 includes the underlying data with extraction and integration processes, section 6 and section 7 explain the approach of multi-view ontology engineering and semantic annotation in detail, section 8 presents the execution result of value chain analysis, which is followed by the summary and future work in section 9.

2 Related Work

The information analysis over semantic integration of public information has attracted significant interest from different Semantic Web communities. The AKT PSI project [4] demonstrated the added-value of richer information insight by using Semantic Web technologies as an important means to integrate a large collection of public information from distributed legacy data sources. The underlying data, in less-standardized formats, was converted into RDF following local data-centric ontology. The different data sources were integrated using an ad-hoc ontology alignment approach.

The EDEN-IW project [12] integrated the water environmental information from heterogeneous databases to support cross-country decision-making processes at the European level. The semantic heterogeneity in terms of different data structures and multi-lingual representations was harmonized at the conceptual level by deploying a multi-layered ontology model. The conceptual viewpoints of local database ontology and user ontology were connected via rule-based mapping to the global ontology. The high-level user queries were translated into global terms and corresponding SQL syntax was constructed and executed over the legacy databases.

The first MBI prototype [14] extracted unstructured and semi-structured text from different web sources, transformed that into RDF and integrated the data with core structured and semantically annotated information. A global ontology was constructed from domain knowledge of the B2B market and was designed to be extensible in order to reuse additional conceptual schema extracted from external sources. The information analysis was performed using specific scenarios such as micro-segmentation and value-chain modeling. These require the application logic to follow either standard business classification schemes or ad-hoc operational logic of individual use cases.

An important goal for the project is to achieve a rich analysis insight over a wide range of data sources. The design of above applications was normally driven by specific key use cases in the closed domain in order to satisfy the mutual interests of particular user group. This could have led to a data-centric approach to save the cost of knowledge engineering. However, the potential diversity of operational logic in real business scenarios cannot be fully addressed in such an approach. The project identified a number of issues:

- The global specification of domain knowledge is too coarse and needs refinement to reflect the requirement of the individual analytical case.
- It can be too expensive to develop a complete global domain ontology that
 meets all requirements. This requires close collaboration between project
 stake-holders including domain experts, knowledge engineers and software
 developers.
- The representation of domain knowledge amongst different user group needs
 is rarely agreed as being common. In the extreme circumstance, the application development may have to repeat the knowledge learning process for
 every case resulting in separate solutions for each partner. The knowledge
 learning can become a very time-consuming process.
- There may be security concerns regarding sensitive information that cannot be shared.

The development process can become extended as new business scenarios emerge and changes to the analysis logic result in significant rework of the software implementation.

3 Background

A Value-Chain is defined by Porter [11] as a series of value-generating activities. Products pass through all activities of the chain in order, and at each activity the product gains some value. In MBI, the value chain is addressed as a network path consisting of relationship fragments between companies. The value-chain application is a semantic search and navigation tool featured with special interest in supporting the modeling of company relationships and the capability to identify key trading behavior and entities that have significant influence on a company's propensity to buy. The relationship is defined at a conceptual level as typed connections between two sub-types of company following specified patterns of trading behavior. In practice, the conceptual model regarding company classification scheme and trading behavior can vary with each individual case and can change frequently. The value-chain model needs to be analyzed in the context of a specific business scenario and user understanding of the industry domain.

4 Application Architecture

To achieve improved use of information the MBI project proposes a multi-view ontology approach to resolve the problem of varying user scenarios. It enables individual user conceptual views over a value-chain network as well as cross-domain knowledge

manipulation by domain expertise. The approach is prototyped within a generic framework architecture, as shown in Fig. 1, which separates the definition or configuration of a form of marketing analysis from the execution or visualization of a specific analysis of the market.

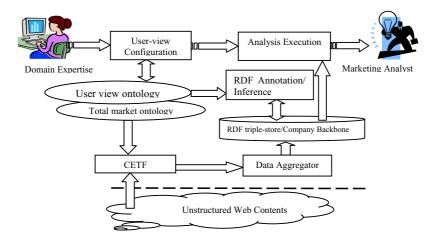


Fig. 1. Generic architecture of value-chain application

The Total Market ontology is a central knowledge base capturing the common conceptualization of a company profile. The common conceptualization includes the general definition of upper concepts and relationships common to all potential analysis projects.

Besides the common concepts and relationships, the refined viewpoint of an individual user scenario is addressed in the User View ontology that extends the Total Market ontology. The information access is enhanced by the augmentation insights of market including both common view and special view.

The user-view configuration is a web-based ontology editor that allows experts of industry domain to define their own knowledge view over the total market. The generated output is a user-view ontology that extends the general class and relationship definition from the Total Market ontology with additional finer definitions.

The Common Extraction and Translation Framework (CETF) is an ontology-driven process that extracts unstructured information from selected external web site. The web sites are of particular interests about trading relationships between companies. The output of web extractions is translated into RDF and follows the semantics of the Total Market ontology or its extensions.

The extracted RDF needs to be integrated with company profile backbone to establish a solid data foundation to support high level queries. The individual view of user conceptualization is not included in the extracted RDF because the extraction process is designed to be generic. The extraction process is completely independent of

the analysis process. The application has an additional RDF annotation process that tags the extracted RDF triples into the specific user view. The details of multi-view ontology engineering and data annotation approach are further explained in Section 6 and Section 7. The marketing analyst is operating over the analysis execution interface where value-chain visualization and navigation is supported over high-level semantic query according to user-view conceptualization.

In this architecture, the knowledge representation is separated from operational process to achieve more flexible information access. The ontology is used here as an extensive knowledge vehicle that allows richer semantic definitions to represent complex analysis scenarios whilst diversity of user-views conceptualization is effectively maintained by introduction of user-defined concept and relationship over Total Market ontology. The concepts and relationships are restricted by semantic rules and mapped to super-classes in the Total Market ontology. The RDF annotation process loads the mapping rules, performs a forward-chained reasoning and generates SPARQL syntax to query the underlying integrated data. The logic consistency and answering correctness between domain expertises and marketing analyst is guaranteed.

Both types of users, the Marketing Analyst and the Domain Expert, interact with the system through a web portal. This portal is developed using the Django framework¹ and acts as a bridge to a collection of backend services. The collection of backend services work tightly with a set of JavaScript widgets in order to provide easy interaction and visualization functions over the semantic data. The list of functionalities is as follows:

- 1. Extract Graph Based Information: this data is extracted by a JavaScript library developed ad-hoc for this project². The first prototype used Prefuse³ but it did not meet the performance. Our implementation for visualizing graphs is Java Applet free and, it uses the HTML Canvas component to render the objects in the browser.
- Represent Tree Based Structures: this service offers to the ontology editor the possibility of querying any graph structure and retrieve it in the shape of a tree. This functionality is flexible enough to allow users the creation of trees from RDF graph.
- 3. Extract Table and List Based Structures: the prototype uses extensively list and tables data structures. This service follows the schema definition for the data sets in order to retrieve the information in such a way as to be properly visualized in the front-end.
- 4. Schema Browsing Service: this service provides a unique point to query the ontology schema using T-Box reasoning. The front-end invokes this service to display a comprehensive view of the structure of RDF data (see Fig. 2).

¹ http://www.djangoproject.com/ (last accessed in April 2009).

² All the graphs-alike screenshots on this paper are examples of the mentioned JavaScript library.

³ http://prefuse.org/ (last accessed in April 2009).



Fig. 2. Schema Browsing Service

5 Underlying Data

The MBI project has aggregated information from heterogeneous sources including structured data, semi-structured data and unstructured data. The company backbone is converted from an industry provided structured data source containing 3.7 million company profiles covering the total UK market. This is augmented with web-based content extracted from selective Web sites.

5.1 Company Backbone

The company backbone is created from an industry provided relational database dump. The company profile data is converted into RDF following concept and relationship definitions in the Total Market ontology. The company backbone is updated on a monthly basis providing an ongoing view of company profiles which can be analyzed over a historical period. In MBI, company backbone data is saved into persistent semantic storage of 4Store⁴ RDF database and each update has more than 160 million triples.

5.2 Ontology-Driven Web Extraction

To exercise the value-chain model, the application prototype is used to create projects that focused on the domain of the Building and Construction (B&C) Industry using data that is regularly extracted from the trade web sites with rich contents about B&C projects and contracts in order to build a network of data depicting the historical trading relationship in the UK B&C industry. Two web sites are selected to support this prototype.

⁴ The home page is at http://4store.org (last access in August 2009).

- Architect Journal Specification is a monthly UK-based magazine, which
 covers the issues of specifying products, and there use in building projects.
 The MBI project currently has extracted contract, supply and product information details including 632 building projects, 392 architects and 4318
 suppliers.
- Barbour ABI is the leading provider of sales leads and construction contract data for the UK market. Their UK-based research team track 100% of all planning applications on every planned building project valued over £100,000 and publish contract update on the daily basis.

The CETF supports the generic ontology-driven solution dealing with unstructured or semi-structured web extractions. The CETF keeps a set of extraction patterns that is matched with the semantic definition of class, instance and properties in the ontology. This may include regular expressions, Gate jape grammar[5], natural languages expressions, concept instances, parsers etc. needed to build a gate pipeline and carry out any data extraction. The framework supports extraction from HTML pages with complex structures, such as nested tables or lists. The string similarity algorithm, for example Levenshtein Distance algorithm [8], is used to compute the similarity of table headers against its matching concepts in the ontology. The WordNet based semantic similarity measure [9] is also deployed. Some generic extraction logic is applied to translate semi-structured tabular data into RDF. For example, if at least one column maps to a concept attribute, instances are generated for each row with corresponding attributes. In the case of nested tables, a concept may be linked to other concepts through object properties. Then a column itself may contain concepts rather than just attributes of the main concept. The extraction task is performed on a scheduled or event-driven basis following the user's instruction. For example, the AJSPEC extraction is performed on the monthly basis and ABI extraction is on the daily basis and driven by any data update.

5.3 Semantic Integration for Companies

The output of a web extraction is identified in a different namespace connecting to the URL of extraction source in order to retain their provenance. The extraction output contains richer semantic information regarding company trading relations that need to be integrated with the company backbone to form the proper value chain definition. The comprehensive description of the approach to semantic integration is not part of this paper. The value-chain application has developed and deployed an ad-hoc integration model with fast mapping functions to match companies identified in web extractions to companies in the backbone data. The company matching can be declared as equivalent if the postcode of registered address of two companies is the same and their company names are closely matched. The Levenshtein Distance algorithm[8] is used here to measure string similarity between two company names. The algorithm output is normalized to the length of string and produces a real number between 0 and 1. The similarity is null if two strings are completely different. The similarity is 1 if two strings are exactly the same. If the similarity exceeds the acceptable threshold, the matching is successful and two companies are tagged with OWL:sameAs relation.

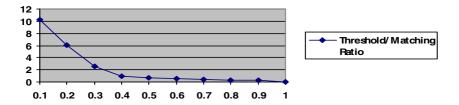


Fig. 3. Company matching using Normalized Levenshtein Distance algorithm

The instance matching result using AJSPEC data based on 4318 suppliers is shown in Fig. 3. Using the matching logic described above, the diagram shows how the matching ratio varies against the value change of similarity threshold. The matching ratio was calculated as the number of equivalent matching divided by total number of companies. The result shows the algorithm ends up with reasonable output if the threshold is between 0.4 and 0.6.

5.4 Data Storage and Query

The persistent storage of the company backbone is supported by 4Store, a large-scale RDF triple-store developed from previous work on the semantic database and query engine, 3Store [6]. The 4Store is capable of scaling up to 60 billion triples⁵ with efficient supports of semantic query via SPARQL [3].

6 Ontology Engineering of Layered Model

The layered model is proposed and deployed in MBI value-chain application to support individual user-view, as shown in Fig. 4. The user-defined concepts and relationship presents the restricted view over the Total Market ontology. The user-defined concept is specified at schema level as classes that are mapped to the Total Market ontology using rdfs:subClassOf relationship plus filtering rules. Semantically, the user-defined company type is a special subset of total market *Organization* featured with unique attribute characteristics. The restriction of such subset are explicitly expressed in filtering rules that lead to a filtering process against *Organization* over underlying data using pre-defined filtering algebra. The user-defined company relationship is also explicitly specified in filtering rules to connect two company sub-types following the specific semantic patterns of trading behaviors in the Total Market ontology.

The data annotation process works as an off-line service to reason the semantic specification of user-defined concept and relationship with filtering rules and find satisfactory evidence from underlying data. The satisfying company individuals are annotated with additional triples in terms of user-defined class and relationship. The relating company profiles, the additional triples generated from annotation process are merged and saved into 4Store as user scenario data. Meta-data about this merging

 $^{^{5}}$ The number was cited from the page, http://www.garlik.com/press.php?id=136-GRLK_PRD

data set is created and linked to the individual user view. Thereafter, the knowledge representation of user conceptualization is transformed into underlying triples and the queries in user-defined terms are supported. The users can easily create new user-views, change and share the contents and make it quickly effective in analysis execution.

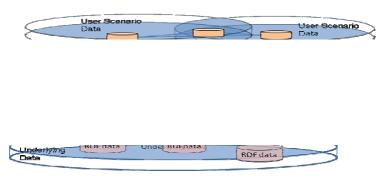


Fig. 4. Semantic adaptation crossing layered Ontology models

6.1 Total Market Ontology

The Total Market ontology was extended from the underlying schema of company backbone following a bottom-up approach, i.e. the ontology schema was derived from underlying database dump with neutral mappings to the conceptual structure of database tables and attributes. The ontology was developed in a generic purpose combining conceptual knowledge from relating domains including people, product, location and transactional information. The ontology has also adopted semantics from other established namespaces including Dublin core[1], OWL-time[7] and SKOS[2]. The ontology model benefits from this method in two extents: Firstly it supports query and reasoning request following the standard classes and attributes definition that would make the application extensible to connect to other applications via an open interface; Secondly, the company profile data is the major part of data aggregation. The straight mapping would save much of transformation efforts to convert relational database dumps into semantic forms.

The conceptualization of total market ontology is shown in Fig. 5, which provides an overall picture over the B2B trading domain with specific focus of *Organization*. The ontology contains many extending branches where further extension is conducted and created in separate ontologies, for examples the offering ontology extends the concept of *Product*, specifying standard taxonomic and part-of relationship in product domain and synonym representations, the business activity ontology describes the company classification in UKSIC Code 2003[10] standard and extends *UK SIC Code*. The extended semantics would allow richer data aggregation of public information and standards from different domains and thus supports finer company classification using comprehensive semantic restrictions.

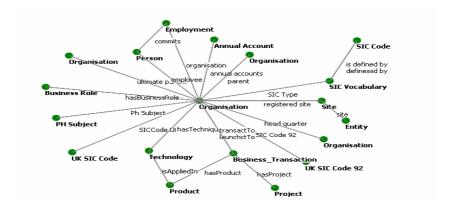


Fig. 5. Domain conceptualization of total market ontology

6.2 User-View Ontology

Information viewpoints model a given representation for some specific point of interest, among the set of possible representations [15]. The user-view ontology offers a particular perspective on the Total Market ontology, capturing explicit specification of additional concepts and relationships required for individual analysis scenarios. In value-chain analysis, the view ontology is developed in a top-down approach, i.e. specifying the user-defined company classification and relationships regardless of the existence of corresponding individuals in the underlying data. The user selects the Total Market ontology or additional conceptual knowledge derived from external Web Extraction to declare the restricted meaning of user-defined concept and relationship. The user-view specification is passed into reasoning service and the underlying RDF is annotated according to user views. The output of annotation service representing the value-chain relationship from different provenances can be further merged or compared to support higher order decision-making processes by the user.

The generic web-based ontology editing interface, user-view configuration, is developed to support user manipulation of view ontology. The underlying syntax of ontology language of OWL and filtering rules is hidden from the user with the ontology presented as a network graph following property domain and range relationships.

6.3 Common Ontology and Project Ontology

To improve the usability of user-view, the ontology representation is sub-typed into hierarchies regarding the common use and knowledge sharing of value-chain analysis. Two types of user-view are created in the prototype application: common view and project view. The common view contains the sharing knowledge, any concept agreed as being common amongst user groups is presented in the common view and to be shared amongst authorized users. The project view includes specific knowledge referring to the individual analysis scenario. It extends the common view ontology and inherits all common definition of company classification and relationship. If one concept or relationship in project view becomes popular and is recognized by most

users, it can be inserted in common view making it visible to other project users. The user can easily create new individual project view from a sharing view of other users. The developing time of user-view conceptualization is saved. The view hierarchies can be further extended to include more layers according to the actual management hierarchy inside the organization, for example sales department and marketing department may share the common set of use scenario whilst the diverse view is still captured in the individual project view. In the value-chain model, the basic common view mainly defines general company classification scheme following UK SIC Code standards, for example, manufacturer, retailer and whole seller. The project includes the additional classification scheme based on the common definition plus semantic restrictions.

6.4 Filtering Rules to Map User-View and Total Market Ontology

The semantic meaning of user-defined concepts is specified as sub-class of total market concept and restricted with filtering rules, which is further interpreted as: the general class in total market ontology is filtered by attribute values of one or multiple connecting concepts following selective semantic path crossing the schema network of total market ontology. For example, a new concept "Manufacturer of Glazing" in individual user-view can be interpreted as: the existence of an *Organization* that has more than 2 historical transactions producing product "Glazing" between "01/01/2009" and "01/05/2009". In such statement, the general concept *Organization* is restricted by the semantic connection relating to concept *Product* and traverse concept *Business_Transaction* filtered by number of transaction and occurrence time period.

The MBI project has developed an ontology-driven, rich semantic interface to guide the user through the specification of such restriction logic and creation of a proper filtering template. The operation of new concept creation is conducted following the steps below,

- 1. The user needs to select the base concept to extend, in the case of valuechain it can be *Organization* or any user-defined sub-class of *Organization*.
- Next, the user is supposed to select a restrictive concept in the Total Market ontology to apply the restriction over base concept, for example *Product*.
 The selection implies the semantics of new concept is limited by existing relations connecting base concept and *Product*.
- 3. The backend ontology service crosses the Total Market ontology graph to find any possible connecting paths between base concept and restrictive concept. The calculation of semantic path considers any class and relationship in the path taking into account of all domain/range properties being explicitly declared with the class or implicitly inherited from its super-classes.
- 4. The user selects one or multiple paths from the preliminary calculation result.
- 5. For each selected path, the user can select to restrict over any class or relation in the path. The setting-up of restriction is guided by pre-defined filtering algebra that works with ontological definitions including data-type property, class definition and object-property.

The preliminary result of ontology computing paths consists of any possible templates regarding the semantic restriction between the base concept and the restrictive concept. The path addresses the unique interpretation of semantic restriction over the base concept following the conceptual specification of the Total Market ontology. Although some paths may not make proper sense to the domain experts, it does comply with the logic structure defined in the underlying ontology. Therefore, the human knowledge of domain expertise is required to validate the preliminary result and make correct choice according to the analysis scenario. The significant flexibility and good data quality is made via combination of formal representation of domain knowledge with diverse understanding of individual scenario from domain expertise. Any multiple selections from the computing paths are regarded as logic AND in the annotation process.

The restriction over user-defined relationship has to follow the same procedure except that relationship is specified as connections between two company sub-types, for example, the user-defined relationship Supply-To may link Manufacturer and Whole Seller following a specific historical transaction pattern. Similarly, the user needs to select correct pattern from computing paths to declare the restriction. The user has options to select from two type of logic pattern, either explicit relationship or implicit relationship. The explicit relationship follows the single path to cross the semantic network and connect two companies. For example, the historical transaction pattern implies that two companies are connected in the relationship if there is at least one trading transaction between them. The implicit pattern defines a flexible model to handle implicit transaction information where interaction between buyer and seller is not explicitly specified. This requires user selection of the third concept to link two company types. For example, user specifies that a Contractor has contactWith connection to an Architect, if two companies have participated into the same building project. The user can also define other types of relationship in a similar pattern, for example the relation competeTo can be defined semantically to connect two companies producing the same type of *Product*. The implicit model can be further extended to cover more complex use cases.

The filtering rules are written in Jena-like[13] rule syntax plus additional extension supporting filtering algebra. Some functions are already supported by the Jena inference engine, the reason of re-definition in filtering algebra is explained in Section 7. The pre-defined filtering algebra includes following functions:

- 1. Data-type property functions:
 - String restriction functions: equals, starts with, contains and ends with;
 - Integer or Float type functions: less than, equals and greater than;
 - Date-time type functions: equals, starts from, ends by;
- 2. Class type restrictions: equals, all value from sub-type, union of;
- 3. Object Property restriction: occurrence (more than), occurrence(less than), occurrence(equals)

The selective combination of the filtering functions and computing paths gives the explicit meaning of user-defined concept and relationship in terms of how satisfied companies and relationships can be derived from underlying data.

7 Data Annotation via Rule-Based Reasoning

The value-chain application needs to query and reason over 200 million triples held in the 4Store. 4Store supports the SPARQL standard without any inference capability. Most inference engines, such as Pellet and Jena, are supposed to work over their native interface to access the underlying triple-stores, i.e. the reasoning engine is bounded with the query engine. To our knowledge, such generic inference engines do not support reasoning tasks by accessing an additional persistent triple-store via SPARQL interface. The MBI project has developed a computational model working as replacement for a generic logic inference engine to reason about filtering rules. The model takes into account of user-defined concept and relationship and filtering rules, to decompose the task, generate SPARQL syntax dynamically, and execute the queries over the underlying data. The underlying data of individual companies are annotated with additional triples generated from reasoning process. The model supports reasoning functions in terms of RDF forward-chained rule syntax to prove the supporting evidence that a company is an instance of classification type or the relationship exists between two companies. These are the primary pieces of information required for a value chain analysis.

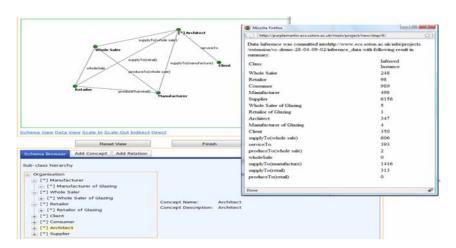


Fig. 6. Semantic reasoning over user-view conceptualization

The example of reasoning outcomes from a user conceptualization is shown in Fig. 6. The result was created from the AJSPEC data in the Building and Construction domain. The connecting network represents the individual user view over the industry domain. The user-defined company types are structured into a type hierarchy tree and the relationship types are named respectively following the historical transactions between two companies. The reasoning is conducted over iterations. The reasoning process tries to fire every satisfying rule in the iteration. The SPARQL syntax is constructed from filtering rules and executed against the underlying data to generate new triples for user-defined concept and relationship. The process continues until no more new triple is generated. The output in terms of generated triples are saved with

Web Extraction data and company profiles creating a new RDF data set, to support queries following the user-view conceptualization.

8 Value-Chain Execution over Project View

The value-chain execution is designed as a generic navigation service to visualize the value-chain relationships between companies according a specific user-view conceptualization. The value-chain model is constructed in a series of steps, each finding the potential connecting companies for a specific company following selected relationships or company types as defined in the user view. The analysis execution does not need to prove the underlying semantic meaning of each concept and relationship. It loads the user-view ontology and composes queries in the high-level user's terms. It executes the query against the user scenario data and visualizes the result in a network graph, see Fig. 7.

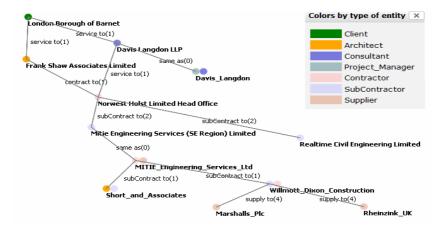


Fig. 7. Value-chain execution and visualization

The example shows the output of a value-chain execution. The construction starts from a client and follows the user-defined relationships to cross the trading network and reach other companies of user-defined types. The relationship closeness is defined by the weight of influential factors such as the number of transactions in the historical transactional data sets. The analysis execution enables quick construction of value-chain models following any user-defined logic view over different industry domains thus providing a generic service required by all project partners.

9 Conclusion and Future Work

The first MBI prototype showed the value of semantic data augmentation from heterogeneous sources. In this paper, we have considered an extension of the approach that covers the more important modeling and sharing of processing logic needed to analyze the underlying data. The extension creates flexibility in support of individual analytical user views to model the reality of specific business scenarios. The multi-view ontology approach was developed and deployed specifically for, but not limited to, value-chain analysis. It extends a single domain ontology to support modeling of varying analysis scenarios and fulfills the user requirements regarding different industry domains. The network analysis is separated at an abstract level into analysis configuration and analysis execution respectively supporting knowledge manipulation of user conceptualization and specific value-chain explorations. The processing logic of the data analysis is persisted in an ontology and aligned with underlying data via an annotation service following generic inference rules. The userview ontology and corresponding annotated data are passed through two sub-phases in parallel, to separate high-level value-chain visualization and navigation functions from the scenario variation of analysis context. The ontology here is used as a super vehicle to convey rich semantic contents of analysis case from domain expertise to marketing analyst, and maintain the semantic consistency throughout annotation process. The information analysis benefits from explicit and extensible knowledge representation using Semantic Web technology. The application logic of analysis scenario is updated in an evolutionary rather revolutionary way by extending and reuse shared ontology views amongst user groups. Much time and effort is saved because user-view conceptualization is simply managed and validated by domain expert via the web-based ontological editing interface. The knowledge transferring is transparent to marketing analyst with less intervention required from software developer, thus performance of analysis project is improved in a long-term scale by using semantic web technology.

The future work is to continue to progress the inference service to provide a comprehensive analysis with more complex semantics, for example the value-chain analysis over an Offering ontology, to allow value chains analysis based on "potential product offering chains" as well as "historical transaction chains".

References

- Dublin Core Metadata Element Set, Version 1.1, http://dublincore.org/documents/2008/01/14/dces/
- SKOS simple Knowledge Organization System Home Page, http://www.w3.org/2004/02/skos/
- 3. SPARQL Query Language for RDF, http://www.w3.org/TR/rdf-sparql-query/
- Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N.R., Tullo, C.: Unlocking the potential of public sector information with semantic web technology. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 708–721. Springer, Heidelberg (2007)
- Cunningham, H., et al.: JAPE: Regular Expressions Over Annotations, http://gate.ac.uk/sale/tao/index.html#x1-1790007
- Harris, S., Gibbins, N.: 3store: Efficient Bulk RDF Storage. In: 1st International Workshop on Practical and Scalable Semantic Systems (PSSS 2003), Sanibel Island, Florida, USA (2003)
- Hobbs, J.R., Pan, F.: Time Ontology in OWL, http://www.w3.org/TR/owl-time/

- 8. Levenshtein, V.: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, vol. 163, pp. 845–848 (1965)
- 9. Lin, D.: An Information-Theoretic Definition of Similarity. In: Proceedings of the 15th International Conference on Machine Learning, pp. 296–304 (1998)
- National-Statistics, U.K.: Standard Industry Classification of Economic Activities pdf (2003),
 - http://www.statistics.gov.uk/methods_quality/sic/downloads/UK_SIC_Vol1.2003.pdf
- 11. Porter, M.E.: Competitive Strategy: Techniques for Analyzing Industries and Competitors. Simon and Schuster (1980)
- Poslad, S., Zuo, L.: An Adaptive Semantic Framework to Support Multiple User Viewpoints over Multiple Databases. In: Advances in Semantic Media Adaptation and Personalization, pp. 261–284 (2008)
- 13. Reynolds, D.: Jena 2 Inference Support, http://jena.sourceforge.net/inference/
- Salvadores, M., Zuo, L., Imtiaz, S.H., Darlington, J., Gibbins, N., Shadbolt, N.R., Dobree, J.: Market blended insight: Modeling propensity to buy with the semantic web. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 777–789. Springer, Heidelberg (2008)
- 15. Spaccapietra, S., Parent, C., Vangenot, C.: GIS databases: From multiscale to multiRepresentation. In: Choueiry, B.Y., Walsh, T. (eds.) SARA 2000. LNCS (LNAI), vol. 1864, pp. 57–70. Springer, Heidelberg (2000)