Purchase Intent, Online Offers and Product Innovation: Misunderstandings in the Ménage à Trois

Davor Meersman[1] and Christophe Debruyne[2]

[1] Digital Ecosystems and Business Intelligence Institute,
Curtin University, De Laeter Way, Technology Park 4,
6102 Bentley, Australia
d.meersman@curtin.edu.au
[2] Semantics Technology and Applications Research Lab,
Vrije Universiteit, Brussel Pleinlaan 2, 1050 Brussel, Belgium
chrdebru@vub.ac.be

Abstract. We discuss a semantic platform that matches a customer's purchase intent against vendor offers. The customers' perception on particular products, including evolving needs and preferences, were captured in a request and product ontology, in turn used to annotate vendor offers. During the project, however, we observed an important gap between the intent descriptions of users and the available data in product descriptions. We argue that through the inclusion of peripheral data, vendors are able to innovate according to customer preference, and users receive increasingly relevant results. We present a method that is essentially a customer-driven innovation system using product innovation ontologies.

Keywords: product ontology, product innovation, online commerce, purchase intent, peripheral data.

1 Introduction

When consumers want to buy a certain item on the Web today, they have to browse through literally hundreds of offers and results and this number is expected to increase in the future. In this model, the vendors drive the process by publishing products and providing means to buy these online. For example travel agencies in the Netherlands need to query many different tour operators to find holiday packages meeting their customers' requirements. They often have an API that facilitates this process, but the granularity of the specific search is often limited due to the heterogeneous nature of all vendor databases.

A solution to this problem would be to allow the consumers to specify their requirements and to match these to offers of different vendors. The COMDRIVE RFP project [5] resulted in a platform enabling consumers to drive the requirements process by expressing their intent to buy a certain product in a tool and language they are comfortable with. This platform sends out the request to a distributed vendor infrastructure, which responds to the request with offers.

For this solution to be effective, a common vocabulary between the consumers and the vendors has to be established. Such a vocabulary can be captured in an ontology. An ontology is commonly defined as: *a [formal,] explicit specification of a [shared] conceptualization* [8]. Ontologies are necessary to enable semantic interoperability between information systems and services on the Web [9]. In general, interoperability is defined as the ability of two or more information systems or their (computerized) components to exchange data, knowledge or resources and to interpret the information in them [4], in this case the COMDRIVE RFP platform and the different vendor applications.

The content of this paper is organized as follows: Section 2 presents the platform and its different components. Section 3 presents the ontology engineering methodology adopted for this project and provides details on the ontology construction processes as well how these ontologies can be used to annotate data in vendor applications. Section 4 evaluates the ontology with respect to goal of the project and points some of the problems we have encountered during the pilot. Section 5 discusses these problems in more depth, arguing the need for a methodology to create a feedback loop from the customer to the producer, presented in Section 6. Throughout this paper, the examples used stem from the domain of winter holiday packages (including winter sports, accommodation, facilities).

2   The Comdrive RFP Platform

In the COMDRIVE RFP project a semantic platform [5] was developed that consists of the following components:

- **Request and Product Ontology.** A semantic and extensible conceptualization of requests and products. An ontology that describes the domain-specific requests, allowing communities of customers and domain experts to contribute product- and domain specific concepts through collaborative ontology engineering.
- **Automated Group Buying Module** allowing community leaders to organize their group buying activities online with their community members with support for member group buying process initiation.
- **Semantic Matching Engine.** Matching of customer intent and vendor offering based on shared, personal customer purchasing profile and community profile for accurate offerings based on predefined and implicit criteria. Matching of semantic product data and flat vendor data annotated with the ontology.
- **Rapid Semantic Node Cloud Navigation.** Dynamic node cloud underpinned by semantic product data for rapid navigation through correlated product concepts. Combines the ease of use of tag clouds with the ability to process structured data. Fig. 1 contains a screenshot of the user interface.

Fig. 2 shows how the different components interact within the platform. Users are able to express their purchase intent through the portal, which has a dynamic interface using the ontology. When requests are entered, the matching engine interprets the request and annotated vendor databases to find matches, which are sent back to the costumer. The whole platform is driven by the request and product ontology and is semantically underpinned in that sense.

Fig. 1. The interface, driven by the ontology, aids the user in expressing their intent. In this example, the user is asked to give one (or more) possibilities for Ski Area ("Skigebied" in Dutch), Auto-completion relies on accessing the data through the application commitment.
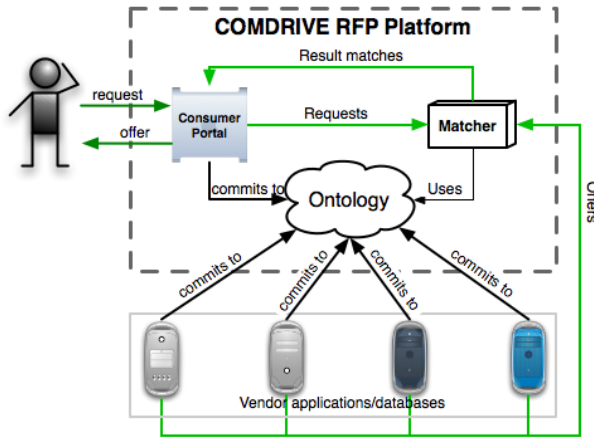

Fig. 2. The interaction between the COMDRIVE RFP Platform's different components

The ontologies do not emerge by themselves, buyers and vendors need to share and reach an agreement on a common vocabulary of the domain. More specifically, software agents need to interpret the information in the customer's purchase intent to automatically match this information with vendor offers based on their semantics. A conceptualization provides a shared agreement on the semantics of core concepts and the relationships between them, imposing a structure on the domain that is readable by both humans and machines.

3   Ontology Engineering Methodology and Related Work

Out of the many collaborative ontology engineering methodologies that exist today [18], we have adopted DOGMA [15], which stands out for its groundings in linguistics. DOGMA relies on the fact that knowledge building blocks, expressed in natural language, are easily obtained and agreed upon (as inspired by database modelling methodologies such as NIAM [21] and ORM [10]), allowing domain experts and knowledge engineers to use natural language to communicate and capture knowledge. The knowledge building blocks - called lexons - in principle only need to express "plausible" facts (as perceived by the community of stakeholders) in order to be entered into the Lexon Base, a repository containing large sets of such lexons. A lexon is formally described as a 5-tuple <G, head, role, co-role, tail>, where G is an abstract context identifier (e.g., a document on the Web) and used to identify unambiguously (to human users at least) the concepts denoted by the term and role labels. Ontologies in DOGMA are selections of such lexons with constraints on their usage (e.g., "A person can have at most one Name").

It should be clear that DOGMA is only the method to reach agreements amongst different stakeholders. Ontologies in DOGMA are actually "representation agnostic": they can be implemented with other formalisms such as RDF(S) and OWL. DOGMA thus precedes the implementation of ontology, and can be repeated to incrementally grow and refine the ontology. Once agreement has been established and an appropriate mapping of (for instance GoodRelations) has been given, the concepts in the DOGMA ontology can be integrated with that particular schema by generating the necessary classes and properties.

Before building an ontology from scratch, one has to assess existing meta-models that describe products (not necessarily implemented with Semantic Web technologies). [12] We have analysed and compared four important product meta-models: eCl@ss[1], UNSPSC[2], EOTD[3] and RosettaNet Technical Dictionary[4]. Both eCl@ss and UNSPSC are broad: the first was created by and driven by the German industry and is thus a "de facto standard", whereas the United Nations Development Programme drives the development of latter. Both UNSPSC and eCl@ss provide very little detail for the travelling domain. The others were designed for more technical industries and did not fit the scope of this project.

Travel industry meta-models include Hi-Touch[5], OnTour[6], Harmonise [6] and the Open Travel Alliance specification[7]. Harmonise focuses on accommodation and events (e.g., sports and conferences), but its main aim is to transfer data between tourism industry partners. Hi-Touch is a commercial thesaurus implemented in OWL to align different vendor databases. OnTour, a recent initiative, mainly covers accommodation and activities. Open Travel Alliance provides a structure for

---

[1] http://www.eclass-online.com/
[2] http://www.unspsc.org/
[3] http://www.eccma.org/
[4] http://www.rosettanet.org/
[5] http://www.mondeca.com/
[6] http://e-tourism.deri.at/ont/index.html
[7] http://www.opentravel.org/

electronic messages, e.g., concerning flights, insurance, etc. Hi-touch and OnTour ontologies were developed based on international standards whereas Open Travel Alliance and Harmonise provide their own.

We bootstrapped the product ontology drawing inspiration from the existing meta-models and vendor applications, which were then refined and completed by several domain experts. We consulted domain experts with different views on the domain, such as tour operators for the vendor perspective and a community of skiers for the buyer perspective. Fig. 3 shows few of the hundreds of lexons created for the purpose of this project.
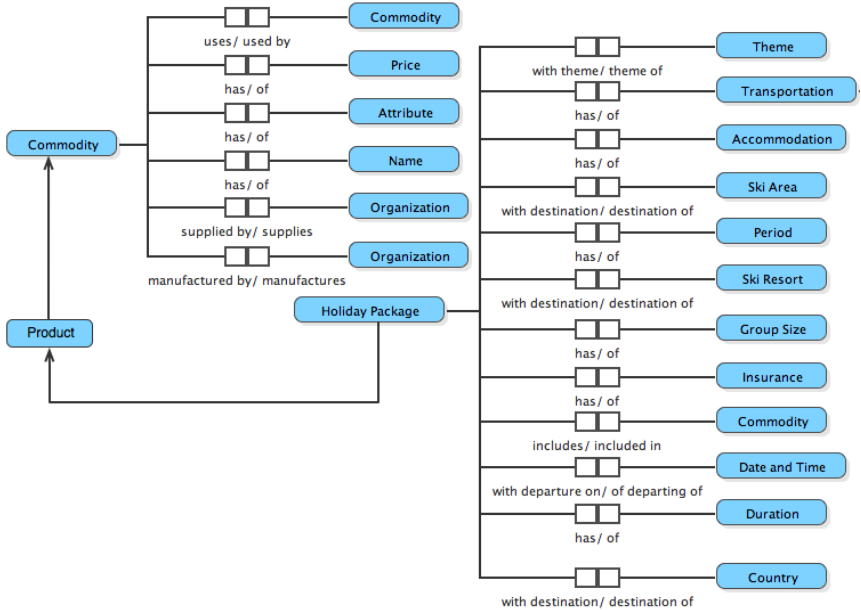

Fig. 3. Some lexons describing domain knowledge developed during the project. These lexons describe holiday packages and commodities, and how holiday packages inherit the properties of commodity by the is-a relationship denoted by the arrow.

The ontology was developed in a modular way. The *Upper Common Ontology* contains the conceptualizations and semantic constraints that are common to and accepted by a general domain, in this case `Product`. For instance, the lexon `<G, Product, with, of, Price>` is true for all applications of stakeholders within that domain and therefore belongs to that layer. The *Domain Application Ontology* contains lexons specific to a certain application domain. In the case of COMDRIVE RFP, these lexons will contain the terms `Holiday Package` and `Accommodation`. The *Lower Common Ontology* represents the interpretation of the domain from the perspective of an organization or community. For instance, the representation of a `Price` might change depending on the community: from a buyer's perspective it is represented by a `Range`, whereas from a Vendor's perspective it is

represented as a `Value`. Whilst the ontology evolves, this layer contains the information that is going to be refined by a core domain expert to be integrated in the Upper Common Ontology. The different modules are then connected by matching context-term pairs. In the lexons shown in Fig. 3, the facts around Holiday Package would belong to the DOA (including the fact that a `Holiday Package` is a `Product`) and the facts and the facts around commodities and products to the UCO.

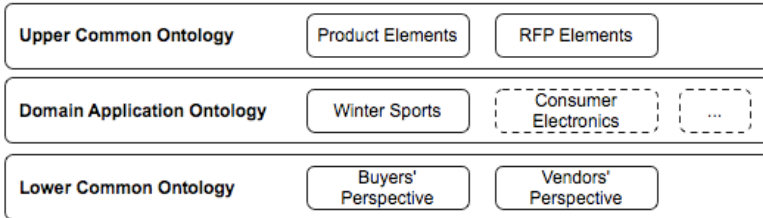|		
Upper Common Ontology	Product Elements	RFP Elements
Domain Application Ontology	Winter Sports	Consumer Electronics	...
Lower Common Ontology	Buyers' Perspective	Vendors' Perspective

Fig. 4. The modular structure of ontologies modelled with DOGMA within the COMDRIVE architecture

Applications commit to the ontology by annotating the application symbols (e.g., fields in a database or tags in an XML schema) with concepts and relations in the ontology. An application commitment thus represents an explicit interpretation of an ontology for an application or a family of applications. In DOGMA, those annotations are made with Ω-RIDL [20]. It consists of a selection of lexons from the ontology, which are relevant for the application, the constraints to specify how that selection can be used (mandatory and uniqueness constraints, for example) and a set of mappings between the application symbols and the symbols used in the ontology. It also provides some scripting functionalities allowing database programmers to manipulate instances whilst accessing the data. Fig. 5 shows some examples of Ω-RIDL statements.

Constraining lexons
`Holiday Package is identified by Name.`
`Holiday Package has exactly 1 Name.`
`Holiday Package has at most 1 Name.`
Mapping application symbols
`Map "/items/item" on Holiday Package.`
`Map "/items/item/title" on Name of Holiday Package.`
`Map "/items/item/description" on Description of Holiday Package.`

Fig. 5. Example of how lexons within an application commitment can be constrained and mapped onto application symbols

This method was applied to construct the ontology that drives the COMDRIVE RFP platform. The matching engine exploits the application mappings expressed in Ω-RIDL to access the data by generating queries in vendor applications for comparison against the intent. As the interface of the platform is also driven by the concepts and relations described in the ontology, the COMDRIVE RFP platform is completely underpinned by semantics.

4   Evaluation

Milq Media[8] is the publisher of the *wintersporters.nl* platform and agreed to take part as pilot partner in our project. Wintersporters.nl's content is characterized by its actuality and a large amount of community generated information. With an average reach of more than 600.000 visitors per month, it is the largest winter sports platform in The Netherlands. The community of wintersporters.nl, with Milq Media acting as the community leader, agreed to be the pilot partner within this project. The pilot ran from Monday 4th October 2010 until Friday 29th October 2010. It was agreed with Milq Media to let the community test the interface to validate the assumptions and results of the project. Their forum provided feedback that enabled us to solve some of the initial bottlenecks in the pilot (e.g., suggest a starting point when users don't know where to start). 38% of the purchase intents were completed.

We presently focus on the pilot evaluation aspects relating to the ontology. These aspects include: result accuracy, result completeness, and concept coverage. The accuracy of results improved significantly during the pilot as we were able to tweak the parameters of the fuzzy matching engine according to the actual user queries to deliver better results.

Because of the heterogeneous nature of the vendor data streams, data that was present with one vendor, was sometimes not present with another. For example pictures were not present in all offers, which has implications for the uniformity of the results sheet.

The most important issue we have experienced was a serious discrepancy between what concepts and level of granularity the user thinks are important in defining an intent (the "ideal image") and what is offered by vendors. Multiple accounts on the pilot forum (which had over 100 posts) related to the inability of users to express their intent fully. For example the size of the ski area, the height of the ski area, the presence of après-ski facilities, the calmness and/or cosiness of the environment, the presence of nice restaurants, the ability to buy travel insurance, ski bus distance, the grade of luxuriousness of the hotel or apartment, etc. This kind of data is not present in the vendor data streams, but constitutes *peripheral data* that is important in booking a ski holiday. Peripheral data is data that is used by the customer in the purchase decision process, but that is not part of the offering of the product or service provider. This kind of data directly relates to the function the user want to see performed by the service they book: to be relaxed, to have fun, to be satiated with good foods, to enjoy the scenery, to be safe, etc.

We have seen that merely building ontologies on top of the data that is provided by vendors does not solve the problem of finding products that match user needs. When given the liberty of defining one's wish, users demonstrate the desire to involve peripheral data that pertains to the product application domain. Although some of these concepts were present in the ontology, there was no data to work with.

Committing to the ontology is not just a question of mapping existing data fields. The commitment will have to entail changes to the internal vendor data structure, for example added granularity or new peripheral concepts. As the business environment changes, vendors need to put on the hat of or work with their producers and innovate

---

[8] `http://www.milq.nl/`

products and their respective representations so that they better correspond with the purchase intent of users.

When a user searched for a product he or she had an ideal image in mind. Through queries and navigating the node cloud the user approximated his or her ideal image by mapping the offers to what was in mind. Eventually the user made a choice and purchase if the offer mapped satisfactorily to his or her ideal image (notwithstanding eventual ideal image transformations influenced by and during the search process). What was completely lost in this process is the original intent, the ideal image or the imagined product the user had. We just know that he or she purchased something that approximated it. This misunderstanding shown in the triangle in Fig. 6 means that no optimisations can be made in terms of product offering, categorisation, presentation or engineering, other than through exterior (and posterior) processes like customer satisfaction analysis and other market research methods.
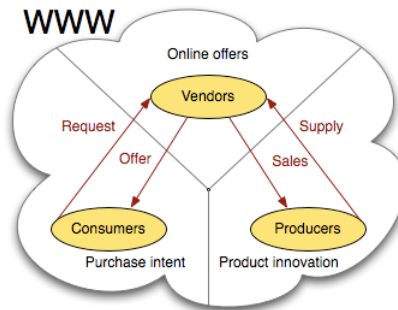

Fig. 6. Producers lack insight in evolving customer preferences and needs on the Web

We would like to argue that for product offers on the Web to improve in terms of matching user needs, the information expressed in purchase intents should correspond to structured data about the product application context. In other words, the entire domain in which the product is used needs to be modelled on the side of the vendor and producer.

In the winter sports pilot, there were several types of peripheral data that were entered by the users: structured data such as ski area size and height, but also concepts such as 'cosiness' and 'luxuriousness', which are user profile dependent.

Integrating peripheral data in one's data structure allows producers to innovate directly based on user desires as perceived through intent definitions and vendors to better tailor their offers to user needs. This is the crux of the problem in product search today: producers fail to realize that users performing queries are a vast (potential) source of direct information for innovation processes.

5 Discussion

Information and knowledge are the most important objects in an innovation process. But it is not clear what information and what knowledge, in what form, at what stage,

in which quantity, at what level of detail is needed. A major problem is that companies generally do not sufficiently grasp the concept of innovation to answer these questions. Innovation is still largely regarded as the result of a bright idea, or as just an organic improvement on an existing product. Generic innovation process models such as for example the Innovanet model [17] do not deal with the intricacies of knowledge, but treat it as a given or as a result. Clearly there are more dimensions to it.

In order to tackle the issue of knowledge intricacy in innovation processes one must:

1. Study and understand the concept of innovation and information in innovation;
2. Develop a conceptual framework to support this knowledge;
3. Embed the framework in a prototype for adaptation and validation purposes.

The type of information that is most important in an innovation process is how the user perceives a product and what it *will do* for him or her. This information is essential in moving the product, but is also exactly the type of information market input in innovation processes should consist of. Producers think in terms of properties or features, whilst consumers think in terms of functions, or the "job they want to get done" by using the product or service [2].
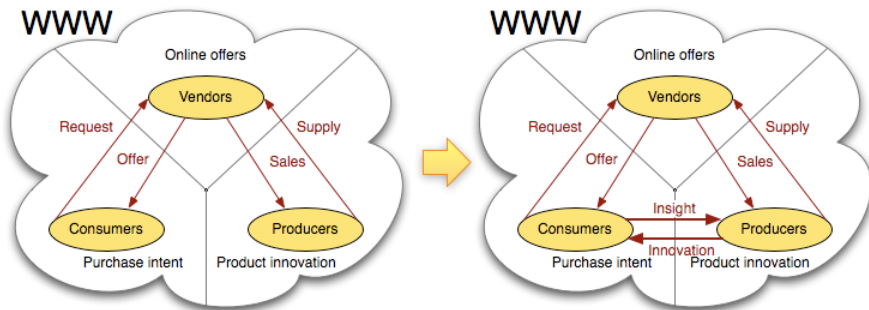

Fig. 7. Producers can use peripheral data to gain consumer insight and drive their product innovation, and (indirectly) supply better information to the customers through the vendors

Whilst the properties of a product are per definition stable and known, the context in which it is used or even how it will be used are usually quite difficult to predict. The notion that Thomas Edison thought of the phonograph as a device that would be used to "record the wishes of old men on their death beds" [7] is only one of many examples that the market frequently puts solutions to use for initially unexpected needs. If an organization wants to keep in touch with all the ways its products are used (and use that information to drive innovation), this information will have to come from the market. Our approach could offer a considerable advancement in bridging the ontology engineering lag. This conceptual dynamics bottleneck is essentially a lack of coverage of concepts by an ontology in a reality that is continually changing. When the concepts in a domain change, there is a "maintenance

lag" in the ontology engineering process [11]. If we look at the causes of conceptual dynamics in corporate environments, these will often be changing market conditions.

The main problem that needs to be addressed is that of a semantic product data approach that is suitable for product innovation purposes. The problem of product data integration has been addressed by the Semantic Web community by developing product ontologies. Apart from the product classification standards discussed in Section 3, the most important product ontologies today are eClassOWL, PRONTO, SWOP and GoodRelations.

- eClassOWL is a product ontology that is derived from the eCl@ss product classification standard that is widely used in the manufacturing industry [11].
- PRONTO is geared towards product information regarding production, storage, sales and distribution [19].
- SWOP, a product ontology developed by the Semantic Web Open engineering Platform project, is an extremely powerful and very granular product modelling ontology for expressing product and component properties [1].
- GoodRelations is a popular lightweight product ontology for representing product information that is relevant to the e-commerce domain [14].

The above ontologies do a good job of representing product, component and related properties. Properties such as product shape, material, colour, family and other manufacturing and sales related concepts are covered, in some ontologies to very great detail. However, in the context of innovation the main shortcoming of these ontologies is that (product, component and property) functions and the application context of the products are not included. Hence they are to be regarded more as very important and valuable assets in the modelling, manufacturing and commerce process, rather than as tools to drive innovation. We argue that for ontologies to be valuable in innovation processes, we need to rework the entire idea of product ontology.

6   Future Work

Based on previous related research [16] in the leisure product domain we will create a product innovation ontology complementing the request and product ontology to include functions and more of the application domain. A major conceptual component of the extended ontology is the integration of product and service functions. Functions answer the question of why something is there. Because the existence of properties is driven by the function they perform, i.e. the fulfilment of initial requirements and in some cases posterior cost considerations, the inclusion of functions and their linkage to product and/or service properties is crucial in the context of ideation.

Another big difference between a regular product ontology and an ontology geared towards product innovation, is that the entire domain of product application is included in the product innovation ontology, as the usage of the product has a great influence on the reasoning behind feature introduction.

The product innovation ontology features a number of concepts for modelling a domain that have been validated in various industry contexts [4]. These concepts are:

- *Actor*: any actor using objects in any process in the domain. The actor is a superset of all possible user profiles. The profiles emerge from the preferences entered by the users. For example younger people tend to put more emphasis on budget and après-ski.
- *Object*: products and components used in processes by actors and objects.
- *Process*: any process in the domain, executed by an actor with the use of objects. The context of a particular process extends different parts of the ontology by introducing additional facts and constraints that support this process. Cfr. Lower Common Ontology in Section 3.
- *Quality*: concepts that define how, to what extent, when etc. something happens; properties and functions of objects, actors and processes. The quality concept also contains concepts pertaining to more intangible customer knowledge such as satisfaction and opinion, characterized by different levels of complexity [3].

For the travel domain we will add more peripheral concepts to the ontology and ensure there is data present for the matching engine to work with. Once the data is there, the effect will be twofold: users will be able to get offers tailored to their intent, and vendors will get insight into what the customer wants.

7   Conclusions

In the COMDRIVE RFP project, the initial outset was to improve product search on the Web by capturing the purchase intent of customers and match that to vendor offers using a request and product ontology and fuzzy matching engine. Although a success in terms of the ontology, we discovered an important gap between the intent descriptions of users and the available data in product descriptions. Users demonstrate the need to use *peripheral data* to describe their purchase intents. This has urged us to revisit the entire idea of product ontology and involve not only the customer and vendor side, but also and importantly the producer side of the equation, to establish a common understanding in the ménage à trois between these economic partners. Through capturing purchase intent including peripheral data (functions and application context), vendors are able to grasp evolving needs and preferences and innovate accordingly. The result is a method that is essentially a customer-driven innovation system that satisfies users with (increasingly) relevant results, and offers vendors a cost-effective way of gathering accurate insight into customer needs and preferences.

References

1. Böhms, M.: The SWOP Semantic Product Modelling Approach. In: SWOP project deliverable 23. SWOP (2006),
   `http://www.swop-project.eu/swop-solutions/semantic-product-modelling-approach/SWOP_D23_WP2_T2300_TNO_2008-04-15_v12.pdf`
   (accessed on June 14, 2010)
2. Christensen, C.M., Anthony, S.D., Berstell, G., Nitterhouse, D.: Finding the Right Job For Your Product. MIT Sloan Management Review 3, 38–47 (2007)
3. Cillo, P.: Market learning capability and innovation. An explorative empirical study in the fashion industry. In: Proceedings of the 3rd International Congress Marketing Trends, Venice (2003)

4. De Leenheer, P.: On Community-based Ontology Evolution. PhD Thesis, Vrije Universiteit Brussel (2008)
5. Debruyne, C., Meersman, D., Baert, M., Hansenne, R.: Community Driven Requests for Proposals: Applying Semantics to match a Customer's Intent to Annotated Vendor Data. In: Proceedings of the 7th International Conference on Web Information Systems and Technologies (to appear)
6. Dell'Erba, M., Fodor, O., Ricci, F., Werthner, H.H.: A solution for data interoperability. In: Monteiro, J.L., Swatman, P.M.C., Valadares Tavares, L. (eds.) I3E. IFIP Conference Proceedings, vol. 233. Kluwer (2002)
7. McDonald, S., Lamberton, D., Mandeville, T. (eds.): The Trouble with Technology. Explorations in the Process of Technological Change. Frances Printer Publishers Limited, London (1983)
8. Gruber, T.R.: Toward principles of the design of ontologies used for knowledge sharing. International Journal of Human and Computer (1995)
9. Guarino, N.: Formal ontology and information systems. In: Int. Conference On Formal Ontology In Information Systems, FOIS 1998, Trento (1998)
10. Halpin, T.: Information Modeling and Relational Databases. Morgan Kaufmann, San Francisco (2008)
11. Hepp, M.: The True Complexity of Product Representation in the Semantic Web. In: Proceedings of the 14th European Conference on Information System (ECIS 2006), Gothenburg, Sweden, pp. 1–12 (2006)
12. Hepp, M., Leukel, J., Schmitz, V.: A quantitative analysis of product categorization standards: content, coverage, and maintenance of eCL@SS, UNSPSC, EOTD, and the RosettaNet technical dictionary. Knowledge Information Systems (2007)
13. Hepp, M.: Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies. IEEE Internet Computing 1, 90–96 (2007)
14. Hepp, M.: GoodRelations: An Ontology for Describing Products and Services Offers on the Web. In: Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management, EKAW, pp. 332–347 (2008)
15. Jarrar, M., Meersman, R.: Ontology engineering – the DOGMA approach. In: Dillon, T.S., Chang, E., Chung, S., Sycara, K. (eds.) Advances in Web Semantics I. LNCS, vol. 4891. Springer, Heidelberg (2008)
16. Meersman, D., Dillon, T.S.: The Open Innovation Paradigm and the Semantic Web: An Ontology for Distributed Product Innovation. In: Proceedings of OTM Workshops 2010 (2010)
17. Paukert, M., Niederée, C., Muscogiuri, C., Bouquet, P., Hemmje, M.: Knowledge in the Innovation Process: An Empirical Study for Validating the Innovation Knowledge Life Cycle. In: Proceedings of the 4th European Conference on Knowledge Management, ECKM, pp. 725–737 (2003)
18. Sure, Y., Staab, S., Studer, R.: Ontology engineering methodology. In: Bernus, P., Blazewicz, J., Schmidt, G.J., Shaw, M.J., Staab, S., Studer, R. (eds.) Handbook on Ontologies, International Handbooks on Information Systems (2009)
19. Vegetti, M., Henning, G.P., Leone, H.P.: Product ontology: definition of an ontology for the complex product modelling domain. In: Proceedings of the Mercosur Congress on Process Systems Engineering (2005), http://www.enpromer2005.eq.ufrj.br/lng/en/index.php (accessed on: May 26, 2010)
20. Verheyden, P., De Bo, J., Meersman, R.: Semantically Unlocking Database Content Through Ontology-Based Mediation. In: Bussler, C.J., Tannen, V., Fundulaki, I. (eds.) SWDB 2004. LNCS, vol. 3372, pp. 109–126. Springer, Heidelberg (2005)
21. Wintraecken, J.: The NIAM Information Analysis Method, Theory and Practice. Kluwer Academic Publishers (1990)