Invoices Case Study

Author: Jose Manuel Gómez-Pérez

Introduction and motivation

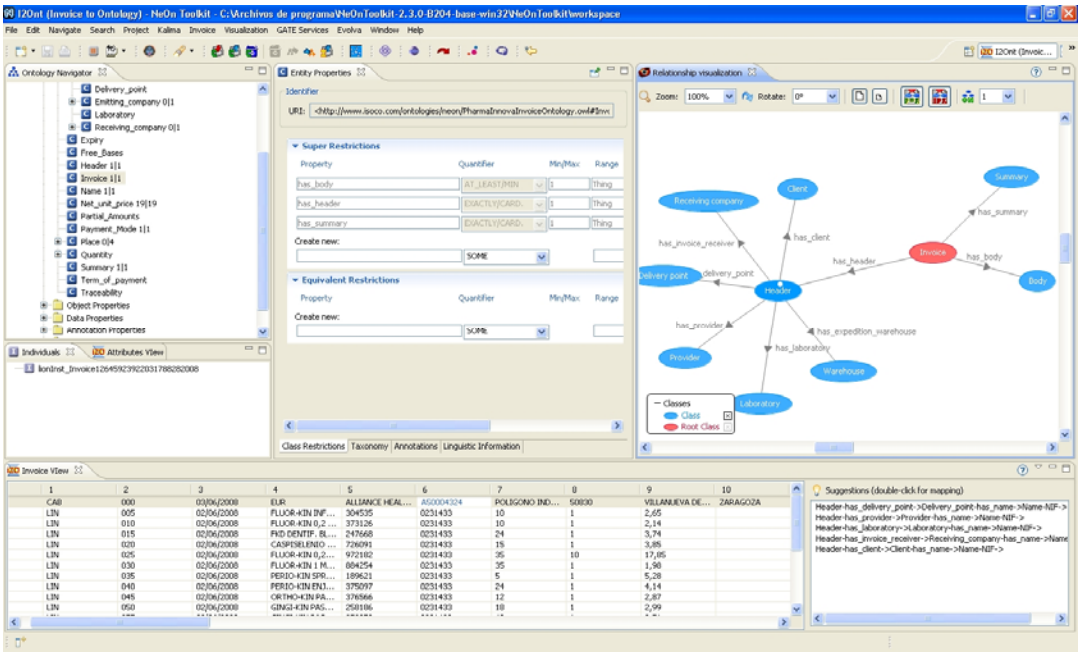
Since the authorization of electronic invoicing by the EU in 2002, the use of electronic invoices for commercial transactions has grown exponentially. Such take up has been accompanied by a large heterogeneity of the means offered by the market to represent and exchange invoice information, as well as by the lack of invoice standards adopted by the main players of the sector.

In this scenario, we have relied on networked ontologies i) to provide a conceptual model of the information related with invoicing, which embraces the different existing standards, ii) to ensure consistency of exchanged invoice data with respect to the formal model of these ontologies, and iii) to allow users to easily define the correspondences between their invoices and a common model, supported by the ontologies in order to automate invoice exchange between business peers.

This chapter deals with the development of such ontologies, in the context of the NeOn electronic invoice management case study, for automating the exchange of electronic invoices in the pharmaceutical sector.

Application scenario (1)

The range of ERP systems managing invoicing information (SAP, ORACLE, PeopleSoft, Baan, Movex, openXpertya, etc.) and the different languages for exchange of electronic business documents that exist in the market (EDIFACT, UBL, IDOC, etc) are extremely diverse. ***I2Ont*** applies the pharmaceutical networked ontologies [1] to enable organizations involved in economic transactions to exchange arbitrary electronic business documents by automatically extracting the information contained in them out of the details of their particular representation formats and technologies, thus saving large amounts of money in the process [2] .



One of the most challenging entry barriers for uptake by real users in the domain, with no background on ontological engineering, is the gap between domain knowledge (e-business and economic transactions in the Pharmaceutical domain) and the formalisms used to acquire and represent such knowledge. Inspired by Newell's definition of "the knowledge level" [3] back in the eighties, we have intended to develop a highly usable, intelligent user interface that enables experts on e-business and financial staff to alleviate their invoice interoperability problems by means of networked ontologies, relieving them from caring about the way invoice knowledge is formally represented, stored, mapped and, in summary, processed. I2Ont allows domain experts to work and think exclusively at the level of their expertise, i.e., electronic invoices.

Invoices Case Study

Application scenario (2)

The solution proposed is grounded on a combination of networked ontologies and a graph-based visualization and navigation paradigm. Networked ontologies provide a formal, semantic backbone between different electronic invoicing formalisms and models, including support for the main invoicing standards like EDIFACT [4] and UBL, and sectorial approaches like PharmaInnova's. The user interface allows for a simple navigation across the relevant invoicing concepts whereas the formal invoice model described in the ontology network allows ensuring correctness and completeness of the correspondence between the different electronic invoice representations.

Previous approaches to the invoice interoperability problem required implementing specific transformations between the formats and models of each pair of organization exchanging electronic invoices. This was cumbersome and little scalable. On the contrary, I2Ont learns by example, i.e., sample electronic invoices are used to define the mappings between electronic invoice data and ontology concepts. Subsequent electronic invoices received by the system, with a format and model compliant with such sample invoices, are transparently imported as instances of the invoicing ontologies by means of applying the mappings defined during the learning phase. From that point on, invoices are automatically exported to whatever invoice format and model known by the system without needing to implement ad hoc (and costly) transformations.

Ontology development life cyle and scenario for building the ontology network

The ontology network life cycle model chosen in the invoicing case study is the iterative-incremental model, given the duration in time of Project NeOn (4 years) and the different degrees of maturity of the required technology. The ontology development was based on a combination of scenario 6 (reusing, merging, and reengineering ontological resources), where we took ontologies like DOLCE Ultra Lite, the W3C time ontology, the TOVE ontology, and the UBL ontology, and scenario 2 (reusing and reengineering non-ontological resoures), where we built on specifications of invoicing standards like EDIFact and UBL.

Ontology requirements (1)

a. *Non-Functional Requirements*
   - The networked ontology for electronic invoicing must be based on eBusiness standards and sectorial approaches.
   - Ontological and non-ontological resources reused in the ontology network, e.g., time and process representations, must be contrasted and approved by the ontological and adopter communities.

b. *Functional Requirements (groups of competency questions)*
   - CQG1. Competency questions regarding the invoicing workflow
   - CQG2. Competency questions regarding multilinguality
   - CQG3. Competency questions regarding inference rules
   - CQG4. Competency questions related to the receiver of invoices
   - CQG5. Competency questions regarding the technology used by the emitters
   - CQG6. Competency questions related to the emitter of invoices
   - CQG7. Competency questions related to time and date management
   - CQG8. Competency questions related to currencies
   - CQG9. Complex competency questions (24 competency questions)

Ontology requirements (2)

Intended End-users. These requirements reflect the views of the following user types:

- U1. User of the invoicing application who is going to model a new invoice
- U2. User who emits invoices
- U3. User who receives invoices
- U4. User who administrates the invoicing system
- U5. Developers of invoicing applications

Ontology processes and activities performed (1)

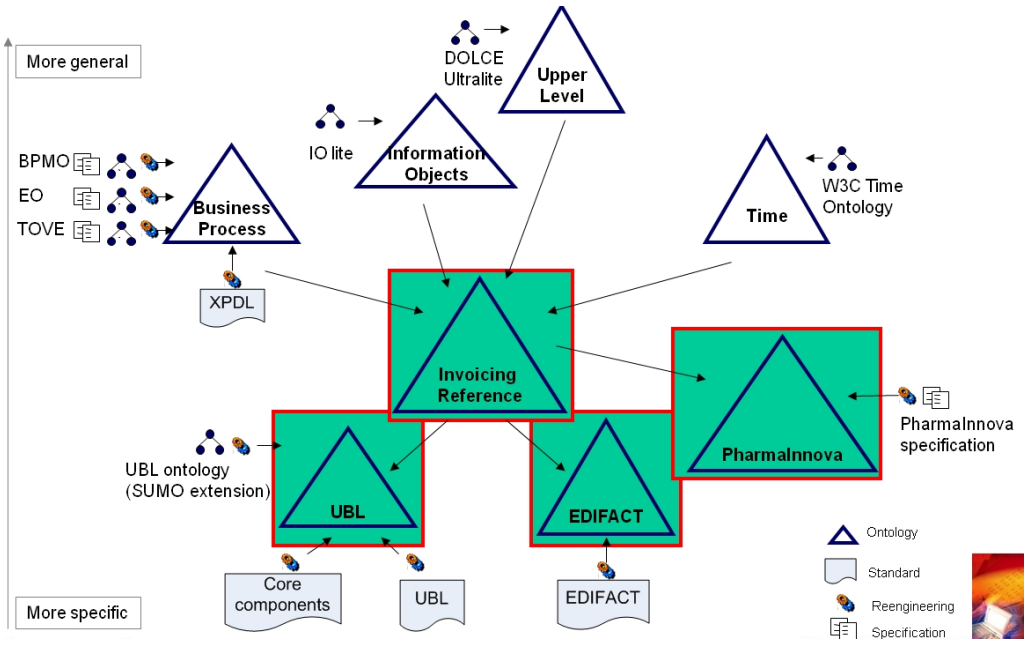
The following processes and activities have been carried out for building the invoice management ontology network:

1. **Ontology Elicitation**. In this activity, the pharmaceutical domain was analysed, with a focus on the invoicing lifecycle, describing the steps an invoice goes through from the time it is emitted to the moment it is validated by the receiving company. This analysis also includes the actors that participate in the process (laboratories, wholesalers and providers), and their requirements.
2. **Ontology Requirements Specification**. In this activity, competency questions reflecting the needs that the ontology had to satisfy in i2Ont have been extracted.
3. **Knowledge Resources Reuse** (Search of existing resources). The resources used for creating the invoicing ontology network can be organized in the following groups:
   - Upper level ontologies and related projects. The motivation for using upper level ontologies comes from the need of reuse of the main reference ontology for invoicing. The purpose of this ontology is to be instantiated for different sectors of the industry. The first instantiation is for the pharmaceutical sector, laboratories mainly, but it will also be extended for providers of these laboratories or wholesalers. These providers provide from chemical products to energy or clean products so they need different instantiations of the invoice reference ontology.
   - Invoicing resources. These resources are mainly technologies for electronic invoicing. The technologies are the Universal Business Language (UBL), EDIFACT, and the PharmaInnova approach.
   - Projects whose main goal is to integrate the invoice vocabulary into ontologies. These include the ONTOLOG project and the XBRL Ontology project.
4. **Ontology Conceptualization** (Development of the invoice ontology network). In this step we conceptualized the resources analysed in the previous activities.
5. **Ontology Specialization** (Adaptation of ontology network). The final invoice reference ontology was adapted to the cluster of companies that were going to use it, a laboratory for instance. The invoice reference ontology will be specialized to each cluster of companies needs (laboratories in an initial phase).
6. **Ontology Localization** (Localization of ontology network). The users of the networked ontology belong to different regions in Spain, in which different languages are used. Spanish is the official language but in these regions there are other co-official languages, therefore localization was taken into account. Likewise, this activity was followed to anticipate future use of the ontology out of Spain.
7. **Ontology Evaluation** (Evaluation of the ontology network). The ontology network has been evaluated by the users of PharmaInnova. Ontology Requirements Specification.

Invoices Case Study

Ontology processes and activities performed (2)

As a result of applying these processes and activities, the following networked ontology was produced for the invoicing case:



Technology used

Both during ontology development and their exploitation in I2Ont, we have used the functionalities provided by the NeOn toolkit together with those stemming from a number of plugins. Among such plugins we highlight the following: the OWL modelling plugin and the Alignment plugin for ontology alignment, RaDON for ontology repair, Ontology RelationShip Visualizer for ontology browsing, CupBoard as ontology store, and SPARQL for ontology querying. Additionally, we have also used GATE to perform named-entity recognition on invoice data.

Additional information:

Contact person: jmgomez@isoco.com

- Gómez-Pérez J.M, Pariente T, Buil-Aranda C, Herrero G. “Ontologies for the Pharmaceutical Case Studies”. NeOn Deliverable D8.3.1, 2007. [1]
- Gómez-Pérez J.M, Pariente T, Daviaud C, Herrero G. “Analysis of the pharma domain and requirements”. NeOn Deliverable D8.1.1, 2006. [2]
- Newell, A. (1982). “The knowledge level”. Artificial Intelligence, 18(1), 87-127. [3]
- UN/EDIFACT Rapporteurs' Team. (1990). “Introduction to UN/EDIFACT with latest news and events”. United Nations Economic Council for Europe. [4]