

## **Part III**

# **Knowledge Elicitation and Management**

Data and information constitute the main building blocks of an information system, a patrimony that must be conveniently exploited to provide the enterprise with the appropriate knowledge, at each decision level and for each business need. This fact is particularly true when the enterprise itself co-operates with fellow-enterprises to amplify its prospects and opportunities.

This part provides an overview on the collection of methodologies, techniques and tools proposed by the ArtDeco project for the purpose of context-aware knowledge elicitation and exploitation in the networked enterprise. Most of these constitute state-of-the-art research, however some more traditional tools are also mentioned, coupled with the most advanced ones, to produce a fully integrated corpus of equipment for: (i) discovery of the useful information sources, in terms of the enterprise documentation, of structured and unstructured data coming from the networked information systems, of web-available knowledge, of sensor data, of event flows within the business processes; (ii) on-the-fly extraction of synthetic knowledge from these information sources, in terms of a common, semantic-based data model represented as an ontology; (iii) dynamic interpretation and integration of the acquired information; (iv) analysis and dissemination of such knowledge to all decisional levels, appropriately adapting it to the user's function and context.

The chapters that follow discuss how both the explicit and the unexpressed knowledge of the various enterprises can come together to enhance the business capabilities of the network in the various phases of the wine production process. Challenges to be faced are the different formats and nature of the available information, from structured data to natural language to sensor-retrieved information, the dynamicity requirement for the data integration process, and the need to spot, dig-out, and analyze the information that is most appropriate for the current operator in the current context of use.

In the first chapter of this part, entitled *Ontology-based Knowledge Elicitation: An Architecture*, Montedoro et al. give a comprehensive overview of the overall architecture of the ArtDeco Web Portal, which, located at the intermediate level (*Shared Information*) of Figure 1.1, provides the following main functionalities: taxonomy-driven word tagging and tag-based querying; knowledge extraction from natural language sources, and concept-based natural language querying; knowledge extraction from applications and sensor networks; ontology extraction from structured data sources; and collection of data from enterprise processes. In this architecture, research prototypes are mixed up with more consolidated tools, yielding a complete framework for the elicitation and exploitation of knowledge. The following chapters give an account of the knowledge-related research that has been carried out in ArtDeco.

The second chapter, named *Knowledge Extraction from Natural Language Processing*, presents a model for knowledge extraction from documents written in natural language. The model relies on a clear distinction between the conceptual level, where the domain knowledge is represented, and the lexical level, which contains the domain vocabulary. The mapping between these two levels is stored as a stochastic model, which takes in account the context of words. The documents' words are appropriately disambiguated during the indexing phase. The engine supports simple

keyword-based queries, as well as *natural language phrase* queries. While extracting and interpreting new terms from the encountered documents, the engine is also able to extend the domain ontology by adding the newly-discovered concepts.

In the chapter *Knowledge Extraction from Events Flows* an analysis of the approaches and methods available for the automated extraction of knowledge from business event flows is presented, focussing on the reconstruction of processes from automatically-generated event logs. Assuming that knowledge can be directly gathered by reconstructing business-process models, various authors in the literature propose that the knowledge acquired be exploited to detect anomalous behaviours in the execution of activities or violations to high level plans. Several of the proposed techniques are briefly examined, and advantages and drawbacks of the different approaches are highlighted, trying to propose a uniform framework for such analysis.

The fourth chapter, entitled *Context-aware Knowledge Querying in a Networked Enterprise*, starts from the idea that the formidable amount of heterogeneous information, accessed by the networked enterprise through all the available channels, makes it difficult for users to find the right information at the right time and at the right level of detail. The research proposes the use of contextual meta-data about the system and the users to reduce this plethora of information, providing high-quality, focussed knowledge to users and applications at all decision-making points. The applicability of the proposed context-aware design methodology and techniques is illustrated within the wine production scenario, where several classes of users access the networked-enterprise data sources, the sensors used for monitoring the productive cycle, and external sources of different nature.

In the ArtDeco scenario data integration of small pieces of heterogeneous information must often be performed on-the-fly, without relying on manual intervention for mapping design; here Semantic Web technologies such as ontologies, which might fail on large-scale data integration, may play an important role. The chapter *On-the-fly and Context-Aware Integration of Heterogeneous Data Sources* describes our ontology-driven framework for dynamic data integration of heterogeneous data sources, where, again, user- and application-queries are dealt with in a context-aware fashion in order to keep the information noise at bay.

Finally, in *Context Support for Analytical Queries* Bolchini et al. propose a solution to analytical query formulation that supports the designer in the OLAP query formulation activity by using the knowledge of the context the decision-maker is currently experiencing. Thus, this last chapter addresses another typical problem of knowledge elicitation, that is, the possibility to provide historical, current, and predictive views of the business operations of the networked enterprise, typical of the Business Intelligence scenario. The major innovative aspect of the proposal is the possibility to connect the relevant contexts of a target application, expressed by using a quite general tree-based dimensional model, with the definition of the analytical queries useful in those contexts; later on, such queries can also be used to select the views that must be transiently or permanently materialized in a data warehouse.