1. Thesis Overview

This thesis presents my contributions in the area of formal representations with applications in e-business. The thesis is focused on two aspects: i) formal modeling of the dynamic aspects of e-business, including business processes and interaction protocols of e-business agents; ii) formal representation and processing of e-business content.

My contributions in the first area concern the development of a formal modeling framework based on Finite State Process Algebra – FSP of e-business agents and processes. This modeling framework was applied to develop a formal semantics of business processes captured as Role Activity Diagrams – RAD. I also proposed a novel modeling language of business processes with formal semantics that was inspired by the Integration Definition for Function Modeling – IDEF framework for function modeling. I show that this language can be integrated with Role Activity Diagrams enabling to capture the functional and dynamic aspects of a business process. These contributions are presented in Chapters 3 and 5.

My results in the second area address the formal representation of e-business content using logic programming. This approach enabled the development of a new class of wrappers for information extraction from HTML documents. The wrappers are able to extract useful information and convert it to a structured format. They are called Logic Wrappers and have a formal semantics based on querying logic programs. Logic Wrappers can be efficiently implemented using XML transformations specified using XSLT. These contributions are presented in Chapter 4.

2. Overview of Chapter 1

In this chapter I introduce the motivation of my research in the area of formal representations in e-business: the challenge of „aligning“ business and formal computational languages. Formal rigor can bring correctness, safety and robustness to the e-business ecosystem. My results can be seen as a contribution towards achieving this general desiderate.

Then I present my own classification of e-business formalisms, focusing on their computational aspects. This classification proposes the following categories: core formalisms (plain text, XML, and JSON), business process modeling formalisms (BPMN, RAD, IDEF, BSDM, and UML), Semantic Web formalisms (RDF(S) and OWL), and formal methods (logics, Petri nets, and model checking).
3. Overview of Chapter 2

This chapter is focused on the results of my research in the area of formal representations of business processes.

In Section 2.1 I introduce the RAD notation and the FSP formal modeling language. FSP is then used to intuitively introduce an example formal model of a sample RAD process.

Then, in Section 2.2 I present my research results regarding the formal modeling of RAD business processes using FSP. RAD diagram is represented as a directed graph. I proposed a mapping algorithm that maps a RAD to a FSP model. This algorithm has a polynomial complexity and this result is stated by Proposition 2.5.

In Section 2.3 I show how FSP models of business processes introduced in Section 2.2 can be formally checked using Fluent Linear Temporal Logic – FLTL.

In Sections 2.4, 2.5 and 2.6, I introduce the Hybrid IDEF formal modeling notation of business processes. In particular, Hybrid IDEF is useful for hierarchical modeling of business processes using the formal concept of decomposition. Decompositions can be „composed” similarly to mathematical functions, thus resulting refined models of a business process. This result is stated in Proposition 2.20. Moreover, a „cut” in the decomposition tree defines an „intermediate” level of detail for presenting a business process model. This result is stated by Proposition 2.23.

Behavioral aspects of Hybrid IDEF business processes can be captured by introducing a mapping algorithm of the Hybrid IDEF notation to Place / Transition nets (P/T nets), a class of Petri nets. This algorithm has a linear complexity with the size of the chosen level of detail of the source Hybrid IDEF business process.

In Section 2.7 I present a method for the consistent integration of Hybrid IDEF and RAD notations into a unified business process modeling notation. The method was implemented as a plugin of the Eclipse platform within the Model Driven Architecture (MDA). The added-value of this proposal is two-fold: i) to bring the possibility of approaching business process modeling from either a mixed functional/dynamic perspective using Hybrid IDEF, or from a more human-like/role-oriented perspective using RAD and ii) to still be able to switch later the perspective depending on the changes in the modeling focus.

4. Overview of Chapter 3

This chapter is focused on formal modeling of e-business agent systems using process algebras. Essentially, in this area I was interested in formal modeling and verification of middle-agents, agent-based service negotiations, and auction systems.

In Section 3.1 I introduce a formal framework for modeling and verification of business agents using FSP and FLTL. In particular, this framework targets the modeling of middle-agents starting from a taxonomy proposed in the research literature. I present detailed FSP models of Matchmaker, Front-agent and Broker agents. The verification part was inspired by formal verification patterns from software engineering.
In Section 3.2 I propose a formal framework based on FSP for modeling service coordination in agent systems. The model was inspired by WS-coordination specification. The approach is demonstrated by a sample model showing how coordination is achieved in a team of agents involved in contracting activities. The model includes the Coordinator, as well as several Contractor and Manager agents. The Coordinator contains several interacting components: ContextManager, ActivationService and ProtocolHandler.

The work reported in Section 3.3 is related to the formal modeling of agent-based auction services. The starting point is the architecture of an agent-based auction server reported in the research literature. The server is configured for running English auctions. The modeling proposed in this section is using the formal framework based on FSP and FLTL introduced in Section 3.2.

Section 3.4 summarizes some other research results that I obtained in the area of e-business agents: distributed collective intelligence, dynamic selection of negotiation protocol, negotiation with concessions, mobile agents, and trust of e-business agents.

5. Overview of Chapter 4

In this chapter I summarize the results of my research in the area of information extraction. My main contribution is the introduction of the new concept of Logic Wrapper, also known as L-wrapper. The main idea of L-Wrappers is to combine the versatility of XML and its associated technologies with the expressivity of logic programming to define a powerful mechanism for information extraction from (semi-)structured documents. In particular this work can be applied to information extraction from Web documents represented in HTML.

In Section 4.2 I review the representation and modeling of information using XML and XSLT. In Section 4.3 I present an overview of knowledge representation and reasoning using logic programming. In particular, I introduce a method for representing and querying structured Web documents using logic programming. This approach naturally evolves to proposing the flat and hierarchical conceptualizations of Web pages for data extraction.

In Section 4.4 I introduce the formal model of L-Wrappers as graphs. An L-Wrapper is defined as a set of pattern graphs that share their extraction signature. Proposition 3.8 states a theoretical result that formulates the necessary conditions for consistent patterns. Then I generalize L-Wrappers to Hierarchical Logic Wrappers, or HL-Wrappers, that are suitable for hierarchical data extraction.

In Section 4.5 I propose an efficient algorithm for automated construction of L-Wrappers based on learning extraction paths from examples. This algorithm has a polynomial complexity for learning bounded extraction paths.

In Section 4.6 I present a method for translating extraction paths to XPath queries. The translation is defined by a mapping algorithm of polynomial complexity. An XPath query can be embedded into an XSLT style sheet to finally extract the information and store it into a database or another structured document. I present example XSLT style sheets resulting from mapping HL-wrappers to XSLT.
6. Overview of Chapter 5

In this chapter I outline my career development plan. In Section 5.1 I present a set of research directions, related to my current research areas that I would like to follow in my future career. In Section 5.2 I focus on my past, present and future teaching activities.

7. References

The thesis contains a list of 119 cited references. This list includes conference papers, journal articles, textbooks, as well as online Web references.