

# SERVICE-ORIENTED SEMI-AUTOMATIC ONTOLOGY BRIDGING

Nuno Silva and João Rocha

GECAD – Knowledge Engineering and Decision Support Group

Institute of Engineering – Polytechnic of Porto

Porto, Portugal

{Nuno.Silva, Joao.Rocha}@dei.isep.ipp.pt

**Abstract** – With the advent of Semantic Web, knowledge-based interoperability faces a new technological shift, in which ontologies and semantic web technologies play a major role. Exploiting the explicit semantic description of the domain of discourse allows reasoning and automatically acquiring semantic relations between two different domains of discourses. Such semantic relations would be further applied in converting data or documents between such domains. This process is named **Ontology Mapping**. This paper proposes a semi-automatic bridging process, based in the idea that while automatic matching algorithms do not provide enough information to overcome semantic heterogeneity, they represent valuable sources of similarities measurements between ontologies' entities, which can be combined into an overall similarity measure according to each service specific requirements.

**Keywords:** *Ontology, Mapping, Architecture*

## I. INTRODUCTION

Many application domains will profit from and strongly require knowledge-based interoperability provided by the Semantic Web approach and technology, instead of the data or information-based interoperability occurring nowadays. In Virtual Enterprises (VE) domain, for example, knowledge-based interoperability is of major importance due to the limitations of data integration in support the dynamic and ill-specified interconnection of entities. In fact, data or information-based integration lacks important features required in VE, like flexibility, interoperability and automatic configuration of interactions between agents engaged in a specific business activity.

Semantic Web, the next WWW trend, suggests the annotation of Web resources with machine-processable metadata, which can provide tools to analyse meaning and semantic relations between documents and their parts. Ontologies as means for conceptualizing and structuring knowledge are seen as the key to the realization of the Semantic Web vision. Ontology allows the explicit specification of a domain of discourse, which permits to access to and reason about an agent knowledge. Semantic Web and ontologies are therefore fully geared as a valuable framework for distinct applications, namely business applications like E-Commerce, B2B and Virtual Enterprises.

However, it is hardly conceivable that a single ontology is applied for all domains and interactions, which raises new interoperability problems. Ontology Mapping intends to solve this problem by defining semantic

relations between ontologies (at conceptual level), and apply those relations at data level over in transforming ontology instances between ontology representations.

The work described in this paper is developed in the scope of MAFRA (Mapping Framework), a conceptual framework [1] covering all phases of the ontology mapping process. MAFRA approach aims to maximize declarativity, complexity, expressiveness and modularization [2]. This paper focuses on the automatic bridging phase of MAFRA process. Semantic bridging is a highly subjective time-consuming task, demanding extensive domain expertise, and therefore even a small automation of the process would be of great usefulness.

In Section 2, the essence of ontology-based interoperability is described. Section 3 describes generically the MAFRA architecture, in which services play a crucial role. Section 4 describes the new automatic bridging methodology. Section 5 will present a short overview of similar or related projects found in literature. This subject is presented at this point in order to argue on benefits and limitations of MAFRA proposals. Finally, Section 6 will provide an overview of the achieved results and point out some current and future efforts. It is considered that the Semantic Bridging Ontology and MAFRA ontology mapping methodology are previously known [3].

## II. ONTOLOGY-BASED INTEROPERABILITY

The ultimate goal of MAFRA is to provide mechanisms to ease and automate knowledge interoperability in agent-based systems, namely concerning Virtual Enterprise and E-Business scenarios. Agents are computer-based entities acting on behalf and according to real world entities goals [4]. Agents are autonomous, reactive, pro-active and socially able. As consequence, agents would require advanced interoperability capabilities, namely concerning communications, interaction languages and knowledge sharing. With the advent and expansion of Semantic Web technologies, ontologies became a more and more frequent approach to describe knowledge in an independent but powerful manner.

Type and extension of heterogeneity between two ontologies depends on different dimensions of the ontologies, specially, described domain, application, granularity [2]. Generally, any ontology is built on three distinct layers: the model layer, the axiomatic layer and lexical layer. The model layer specifies domain and/or application entities, their inter-relations (e.g. subClas-

sOf) and properties. The axiomatic layer constraints the interpretation and application of entities through axioms or rules (e.g. parents of an instance of Person are instances of Person). The lexical layer characterizes entities and their properties with natural language lexicons, giving them a real word meaning (e.g. XPTO entity corresponds to real world entity Person or Individual). Besides the obvious benefits in ontology development and evolution, these three layers separation provide independent elements to reason upon, allowing the argumentation and determination of the extension and causes of heterogeneity.

### III. SERVICE-ORIENTED APPROACH

One of the ultimate innovations in MAFRA is its service-centric approach, which will be described in this section (Figure 1).

This open architecture is based in the notion of service, which is an external pluggable module that provides certain competencies to the ontology mapping system's core processes [1]. Services capture and represent the competencies in dealing with certain type of semantic relation. The primary goal of service is the transformation of the source ontology's instances into

target ontology's instances. However, other competencies in the ontology mapping process are envisaged. One of these competencies is the automatic bridging of entities. Bridging corresponds to the definition, at conceptual level, of semantic relations between source and target ontologies' entities. This is an intrinsically subjective process [5], but even a small automation of the process has great impact in the human intensive, time consuming mapping process.

Ontology mapping aims to define semantic relations between source ontology entities and target ontology entities, which will be further applied at instance level to transform source instances into target instances.

Semantic relations are defined through the instantiation of semantic bridges:

$$\text{semanticBridge}(\text{TR}, \text{SE}, \text{TE}, \text{SC})$$

where TR is the transformation service to be applied in transforming instances of the source entities into instances of the target entities, SE is a subset of source ontology entities that semantically relate to TE, the subset of target ontology entities. Finally, SC is the set of condition expressions constraining the execution of the semantic bridge.

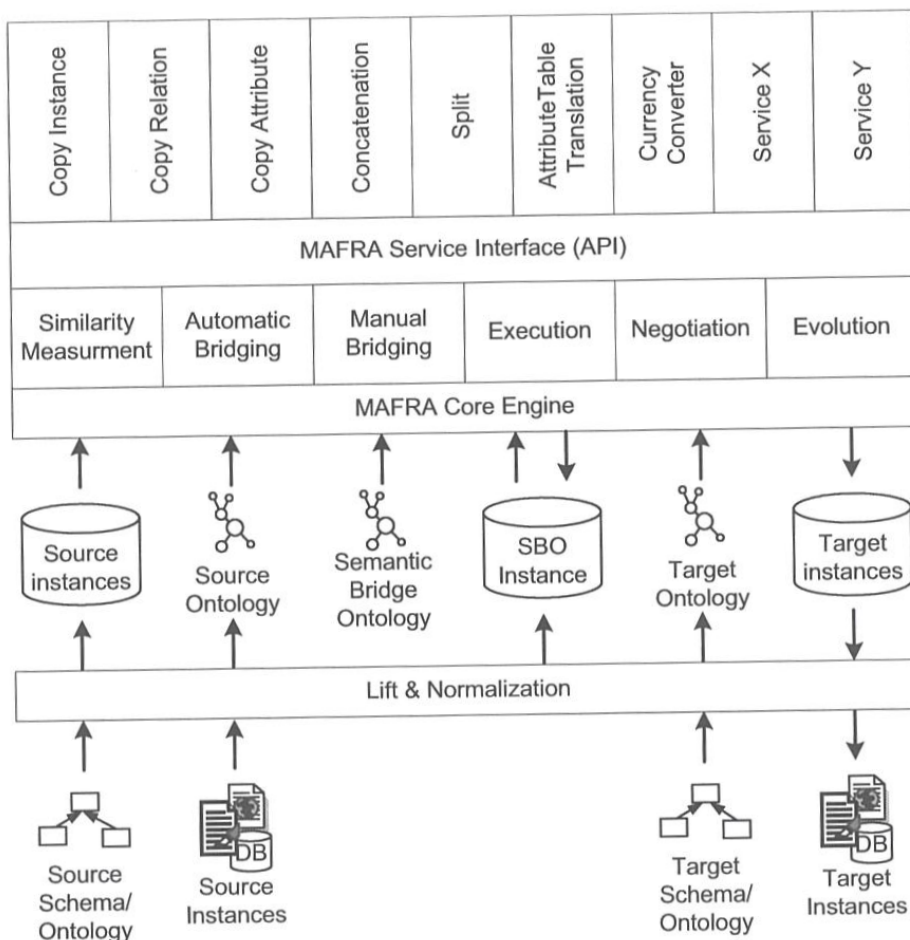


Figure 1- MAFRA System Architecture

**Table I-** Example of standard Services available in MAFRA Toolkit

CopyInstance	Creates target concept instances for each source concept instance that fulfills specified conditions.	
Argument ID	Argument Type	Comment
Source Concept	Concept	Source ontology class whose instances will be transformed
Target Concept	Concept	Target ontology class to create
Extensional Specification	ArrayOfConditions	Extensional definition of source class instances
Generic Conditions	ArrayOfConditions	Constraint of the bridge execution
Minimum Cardinality	Integer	The minimum number of instances to translate
Maximum Cardinality	Integer	The maximum number of instances to translate

Transformations requirements vary enormously from mapping to mapping and from semantic bridge to semantic bridge. Therefore, it is virtually impossible to provide all possible transformation requirements in a centralized static ontology mapping system. This observation lead to the adoption of a modular, decentralized approach, where independent transformation modules are attached to the system functional core modules (e.g. bridging, execution, negotiation, evolution). These modules are called Services and comply with a specific interface, acting as intermediary between services and functional core modules.

Besides the variety of transformation requirements, it is possible to distinguish between two types of transformations according to the type of the target entity [2]:

- When the resulting instance is a concept instance, a new instance will be created and uniquely identified. A concept instance is a container for property instances;
- When the resulting instance is a property instance than the new property instance will always be attached to a previously created concept instance.

Pragmatically this distinction leads to identification of two types of semantic bridges:

- ConceptBridges, those who create target instances, and always apply the CopyInstance transformation procedure. The CopyInstance service is universal, in the sense that all characteristics of the transformation are known *a priori* and are available in the service;
- PropertyBridges, in which the service varies depending on specific transformation requirements.

Each Service is uniquely identified and is characterized by a set of arguments, defining its interface. In the current implementation of MAFRA several transformation services are available, providing sufficient transformation capabilities for a wide range of cases. Table 1 presents the CopyInstance service interface. Semantic bridges entities are connected to service arguments through a map of argument identifiers and values. For example:

```
CB0=semanticBridge( CopyInstance,
    {sourceConcept=<O1>C0},
    {targetConcept=<O2>C0},
    {})
```

While the MAFRA based approach strongly minimizes the mapping process between two domains of knowledge, the process is still very time-consuming and domain-expertise dependent.

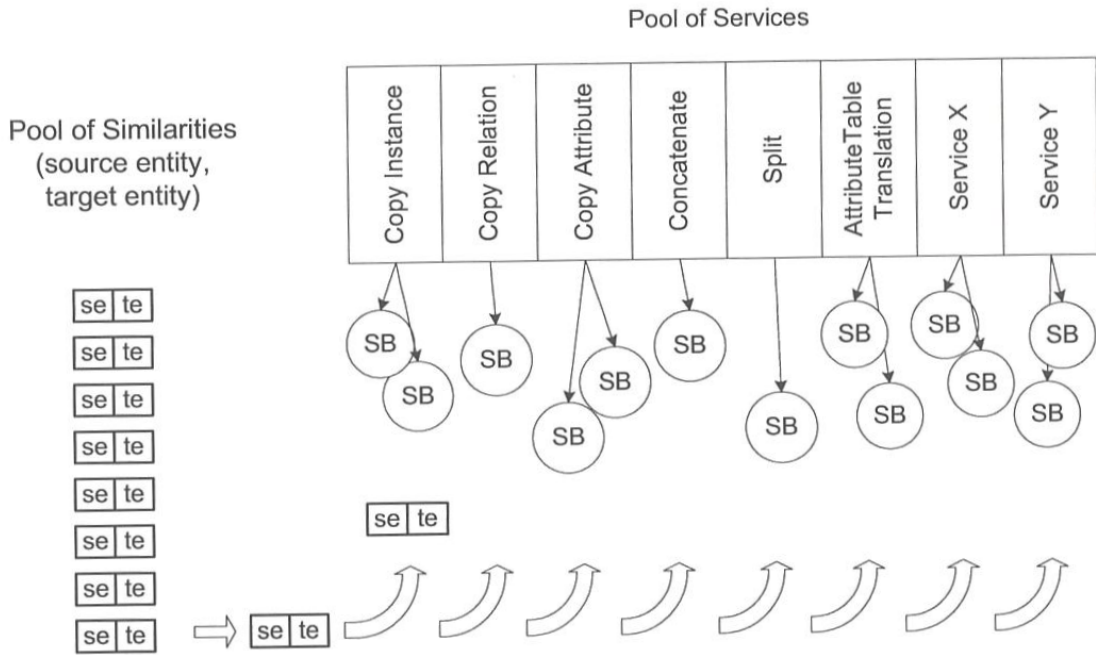
#### IV. AUTOMATIC BRIDGING

Besides the fact the process is inherently subject, which will require the domain expertise supervision and agreement, semi-automatic bridging is nevertheless a step further in direction to complete automatic knowledge-based interoperability. Semantic bridging is a highly subjective time-consuming task, demanding extensive domain expertise. MAFRA provides support and automation facilities. Early stages in the semantic bridging phase typically aim to capture lexical similarities between source and target entities. Such task is very subjective since domain expert deals with natural language description of entities. Recurring to lexical tools, like dictionaries, specific domain thesaurus and WordNet, it is possible to multi-classify similarities between pair of entities.

These classifications are then combined into an overall similarity measure. The result of these early steps is a set of pairs in the form of <Source ontology entity, Target ontology entity>, denoting semantic equivalence between source and target ontology entities. According to a pre-defined threshold level, the similarity pair is kept or disregarded. However, the same entity (source or target ontology entity) might be present in several pair forming therefore a complex n:m map.

Typically, when the problem is too complex, one of the possible solutions is the division of the problem into smaller problems. Such division and modularization is already present in the system (independent modular services), and it seems reasonable to exploit it.

Hence, the proposed process consist in pushing each and all similarity pairs to each and all available services (Figure 2), which in turn determine if the similarity pair:



**Figure 3-** Figurative representation of the automatic bridging process

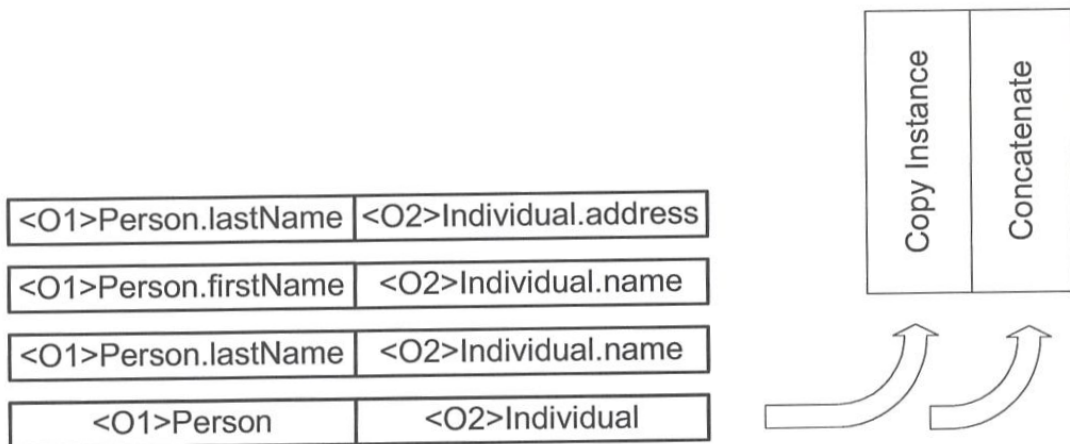
should be added to an already existent semantic bridge, in which case entities are attached to the respective arguments;

is relevant for the creation of a new semantic bridge, in which case a new provisory semantic bridge is created with corresponding service, and entities attached to the respective arguments;

invalidates an already existent semantic bridge, in which case the provisory semantic bridge is deleted.

In order to clarify the proposed methodology, consider the example represented in Figure 3, in which four similarity pairs and two services are represented. To be attached to a semantic bridge, the Copy Instance service requires a similarity pair of concepts. The Concatenate service requires a set (cardinality greater than 1) of similarity pairs, in which the target entity is the same in each and all similarity pairs. The semantic bridge becomes invalid if the target entity is similar to a source entity which in turn is similar to more than one target

entity. Source and target entities must be attributes (literals). In this example, the first pair (lowest pair in the heap) will give rise to a provisory semantic bridge with Copy Instance service (CB1). Second similarity pair is irrelevant for Copy Instance service, since the type of entities required by service do not match. Besides, second pair is accepted by the Concatenate service in order to create a provisory semantic bridge (PB1). The third pair is also accepted by Concatenate in order to complement PB1. Like both previous pairs, fourth pair is refused by Copy Instance service. Concatenate service on the contrary considers it, since both entities are attributes. However, the pair's source entity is already mapped in PB1 with different target entity, which violates one of the Concatenate service constraints. In such circumstances, the Concatenate service becomes invalid and PB1 bridge is deleted. However, lets consider that fourth similarity pair was disregard, and PB1 is valid. The resulting semantic bridges would



**Figure 2-** Dispatching similarity pairs to each service in the system

be:

```
CB1=semanticBridge( CopyInstance,
  {sourceConcept=<O1>Person},
  {targetConcept=<O2>Individual},
  {})
```

```
PB1=semanticBridge( Concatenate,
  {sourceAttributes=list(
    <O1>Person.firstName,
    <O1>Person.lastName) },
  {targetAttribute=
    <O2>Individual.name}, {})
```

Finally, it is necessary to inter-relate semantic bridges. This process is named inter-bridging and aims to find a concept bridge whose concepts are root of all properties in a property bridge. If no concept bridge is found, an inferred concept bridge is proposed to the domain expert between root concepts. In this example, root concepts (<O1>Person and <O2>Individual) are already bridged, and therefore the scope is set:

```
hasBridge(CB1, PB1)
```

The Automatic Bridging process is currently being implemented and first tests are running.

More than in any other development phase, long and strong efforts are expected in the parameterization and customisation of services, and in their interrelation with type of similarity measures.

V. COMPARISON TO RELATED WORKS

Three distinct ontology mapping projects should be pointed as related to the work presented in this paper.

In [6] an extension to Protégé has been described, consisting of a definition of the mapping between domain ontologies and problem solving methods. This approach defines a valuable set of desiderata and mapping dimensions, but lacks some important features, especially in allowing mapping between multiple concepts. Nevertheless, there is no record of experiments that apply it to the Semantic Web environment. The second approach is RDFT [7], a meta-ontology that describes Equivalence and Versioning relations between either an XML DTD or RDFS document and another XML DTD or RDFS document.

An RDFT instantiation describes the semantic relations between source and target documents, which will be further applied in the transformation of documents. Thirdly, the Buster project [8] applies information integration to the GIS domain. Two distinct approaches were proposed: rule-based transformation and re-classification.

The rule-based approach applies a procedural transformation to instance properties, while classification applies class membership conditions to infer target classification through description-logic tools. However, these two approaches are not integrated, which limits mapping capabilities.

In general MAFRA capabilities go beyond those provided by prior projects [2]. In fact, MAFRA is the first approach to integrate all phases and dimensions of the

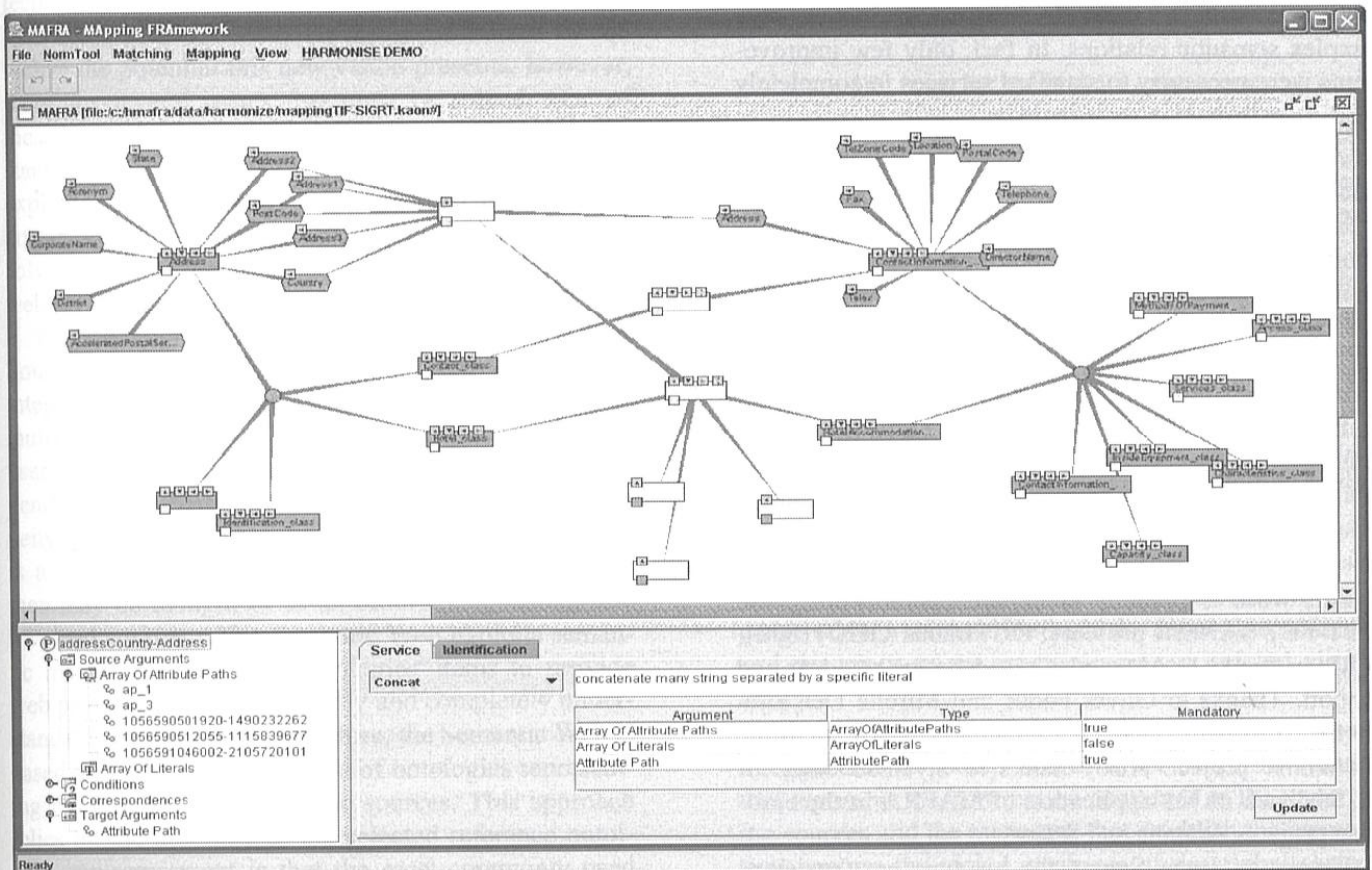


Figure 4 - Screen shot of the Graphical User Interface of MAFRA Toolkit

ontology mapping process into an overall perspective. Semantic Bridge Ontology is, like RDFT, a meta-ontology of the ontology mapping domain, but unlike RDFT, MAFRA allows an unlimited a wider range of semantic relations, instead of the 1:1 (copy) .semantic relation supported by RDFT.

The Graphical User interface and automatic bridging methodology are not mentioned in any of the prior projects. However, MAFRA lacks substantial application and integration experience and do not currently support ontology evolution.

## VI. CONCLUSION AND FUTURE WORK

MAFRA tackles very complex ontology mapping problems with its standard transformation services, but its architecture allows for an easy integration of transformation services, which proves to be important on the semantic web's most common applications: e-business, virtual enterprises and information retrieval. The MAFRA Toolkit implements the ideas described in this paper, providing domain expert with an intuitive, easy to use and integrated GUI (Figure 4). MAFRA was adopted as the representation and transformation engine core technology for the Harmonise project. Harmonise uses an "Interoperability Minimum Harmonisation Ontology" (IMHO) as *lingua franca* between agents. MAFRA is responsible for the acquisition, representation and execution of the ontology mapping between each agent specific ontology and IMHO.

Application with real world ontologies and instances confirmed MAFRA natural skills to deal with very complex semantic relations. In fact, only few improvements were necessary to standard services to completely fulfill transformation requirements.

Our efforts are currently focused in improving the automatic bridging process, specially considering the case where several semantic bridge are suggested for the same similarity pair. We are also interested in the evolution of the ontologies and its consequences to the ontology mapping process. A longer-term project should facilitate the mapping acquisition between different agents using meaning negotiation. Our longer-term research concentrates in negotiation of semantic relations using argumentation between autonomous entities participating in the interoperability.

## VII. ACKNOWLEDGMENTS

This work is partially supported by the Portuguese MCT-FCT SANSKI project (POCTI/2001/GES/41830).

Many thanks to Alexander Maedche for his ideas and support. Thanks to Oliver Fodor and Mirella Dell'Erba for their feedback on the application of MAFRA in the Harmonise project. Also, thanks to Siyamed Sinir for his feedback on the application of MAFRA in the european projects Artemis and Satine.

Thanks to Jorge Santos for his opinions, revisions and sometimes fruitful but always interesting discussions.

## REFERENCES

- [1] Maedche, A. et al., 2002; MAFRA - A MAPPING FRAMEWORK FOR DISTRIBUTED ONTOLOGIES; LNCS, 2473, 235-250; Sigüenza, Spain; 2002.
- [2] Silva, N. and Rocha, J., 2003; Semantic Web Complex Ontology Mapping, 82-88; Halifax, Canada; 2003.
- [3] Silva, N. and Rocha, J., 2003; MAFRA - An Ontology MAPPING FRAMEWORK FOR THE SEMANTIC WEB; Colorado Springs (CO), USA; 2003.
- [4] Jennings, N. and Wooldridge, M., 1996; Software Agents, 17-20; 1996.
- [5] Madhavan, J., Bernstein, P. A., and Rahm, E., 2001; Generic Schema Matching with Cupid, 49-58; Rome, Italy; 2001.
- [6] Park, J. Y., Gennari, J. H., and Musen, M. A., 1998; Mappings for Reuse in Knowledge-based Systems; Banff, Canada; 1998.
- [7] Omelayenko, B., 2002; RDFT: A Mapping Meta-Ontology for Business Integration; LNCS, 76-83; Lyon, France; 2002.
- [8] Stuckenschmidt, H. and Wache, H., 2000; Context Modeling and Transformation for Semantic Interoperability, 115-126; Berlin, Germany; 2000.