A negotiation model for composing ontology matching approaches

Cássia Trojahn dos Santos[1], Márcia Cristina Moraes[2], Paulo Quaresma[1], Renata Vieira[3]

[1]Departamento de Informática, Universidade de Évora, Portugal

[2]Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Brazil

[3]Pós-Graduação em Computação Aplicada, Universidade do Vale do Rio dos Sinos, Brazil

cassia@di.uevora.pt, mmoraes@pucrs.br, pq@di.uevora.pt, renatav@unisinos.br

Abstract. *In this paper, we review previous work on classification schema matching approaches and present an automatic composite solution for the matching problem based on agent negotiation. We consider that different agents working on the basis of particular approaches arrive to distinct matching results that must be shared, compared, chosen and agreed.*

1. Introduction

Ontology matching is the process of determining correspondences between entities from different ontologies. These correspondences can be used for various tasks, such as ontology merging, query answering, data translation, or for navigation on the Semantic Web.

There are different approaches to the matching problem. In this paper we review previous work on classifying schema matching approaches. We believe that approaches developed for schema and ontology matching refer to the same problem. We propose a new classification based on the previous classifications.

Moreover, given the nature of the problem and the variety of proposed solutions, we present an agent negotiation model for ontology matching. We consider that different agents working on the basis of particular approaches arrive to distinct mapping results that must be shared, compared, chosen and agreed. Therefore we propose a negotiation model for agents to negotiate on a final mapping that should reflect a better solution.

This paper is structured as follows. Section 2 presents different approaches to ontology mapping. Section 3 describes the proposed model for ontology mapping negotiation. Section 4 presents the final remarks.

2. Ontology Mapping Approaches

The previous work of (Rahm & Bernstein, 2001), (Shvaiko, 2004) and (Shvaiko & Euzenat, 2004) present a good idea of the various approaches on automated ontology matching. First, we review the taxonomy presented by these authors and then, we present our classification.

2.1 Classification of Rahm and Bernstein

(Rahm & Bernstein, 2001) distinguish between individual and combining matchers (Figure 1). Individual matchers compute a mapping based on a single matching criterion. Combining matchers combine individual matchers in a hybrid matcher or combine multiple match results in a composite matcher.

The individual matchers comprise:

- Schema-based or instance-based level: match can consider instance data or only schema-level information;
- Element or structure level: match can be performed for individual schema elements (e.g., attributes), or for combinations of elements (e.g., schema structure);
- Linguistic and constrained-based techniques: matcher can use a linguistic based approach (e.g., equality of names, equality of canonical name, synonymous, hypernym, similarity based on common string) or a constraint-based approach (e.g., based on keys, cardinalities, and relationship types).

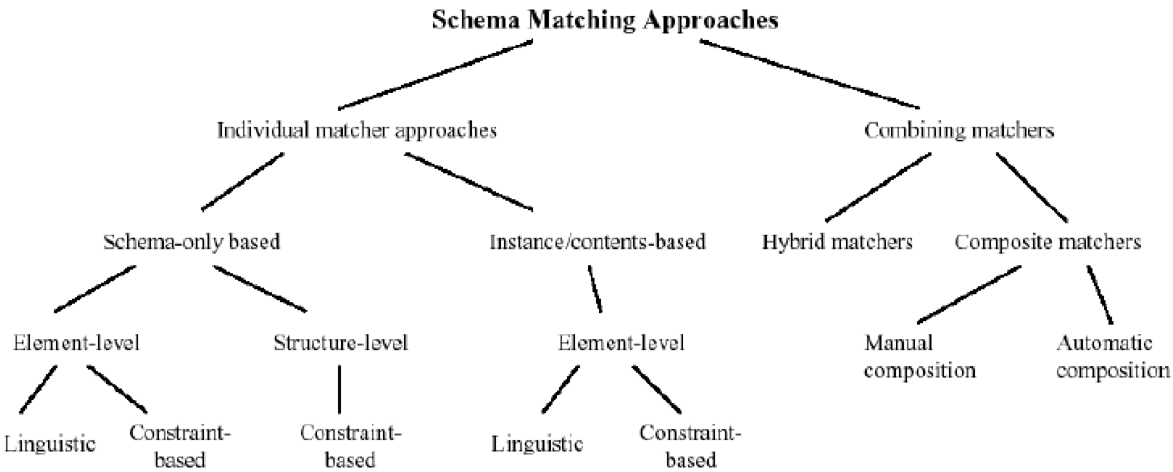


Figure 1. Classification of Rham & Bernstein.

2.2 Classification of Shvaiko

(Shvaiko, 2004) distinguishes between heuristic and formal techniques at schema-level; and implicit and explicit techniques at element-and structure-level (Figure 2).

- Heuristic or formal: the characteristic of the heuristic techniques is that they try to guess relations which may hold between similar labels or graph structures; while the formal techniques have model-theoretic semantics which is used to justify their results
- Implicit or explicit: implicit techniques are syntax driven techniques (i.e., techniques which consider labels as strings, or analyze data types, or soundex of schema/ontology elements); explicit techniques exploit the semantics of labels (i.e., techniques are based on the use of tools, which explicitly codify semantic information, such as thesauruses).

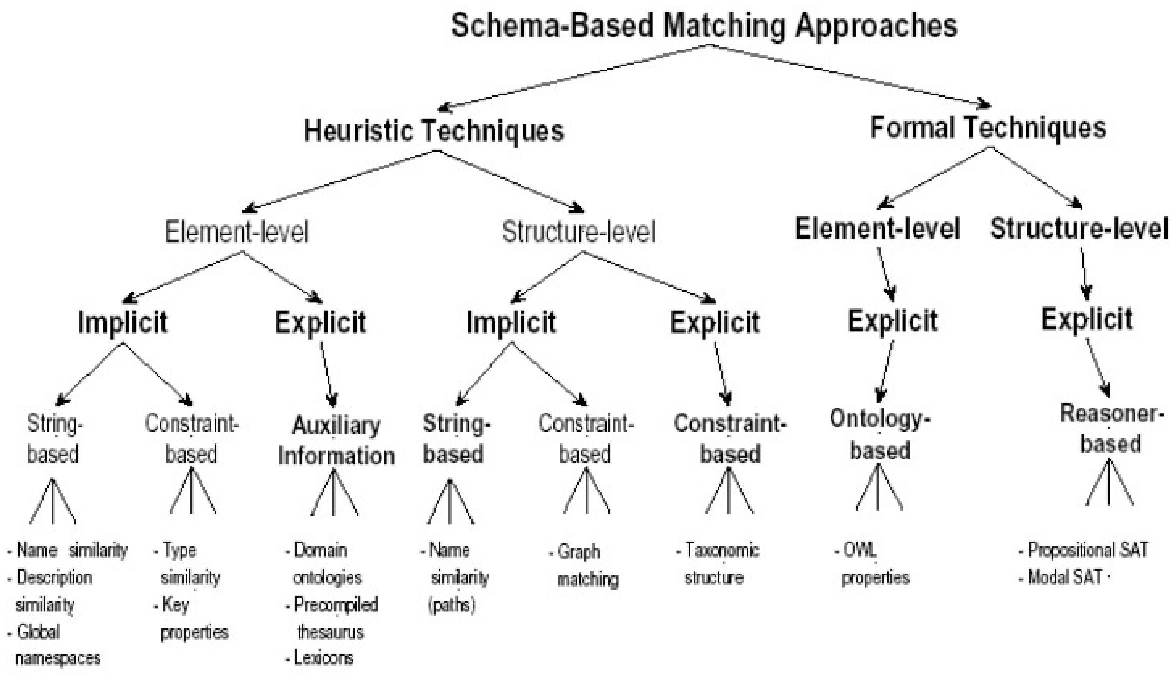


Figure 2. Classification of Shvaiko.

2.3 Classification of Shvaiko and Euzenat

(Shvaiko & Euzenat, 2004) introduce a new criteria (Figure 3) which is based on (i) general properties of matching techniques, i.e., approximate and exact techniques; (ii) interpretation of input information, i.e., syntactic, external, and semantic techniques at element and structure levels; and (iii) the kind of input information, i.e., terminological, structural, and semantic techniques.

2.4 Classification of Giunchiglia and Shvaiko

(Giunchiglia & Shvaiko, 2004) distinguish between weak semantics and strong semantics element-level techniques. Weak semantics techniques are syntax-driven techniques; examples are techniques which consider labels as strings, or analyze data types, or soundex of schema elements. Strong semantics techniques exploit, at the element level, the semantics of labels. These techniques are based on the use of tools which explicitly codify semantic information, e.g. thesauruses.

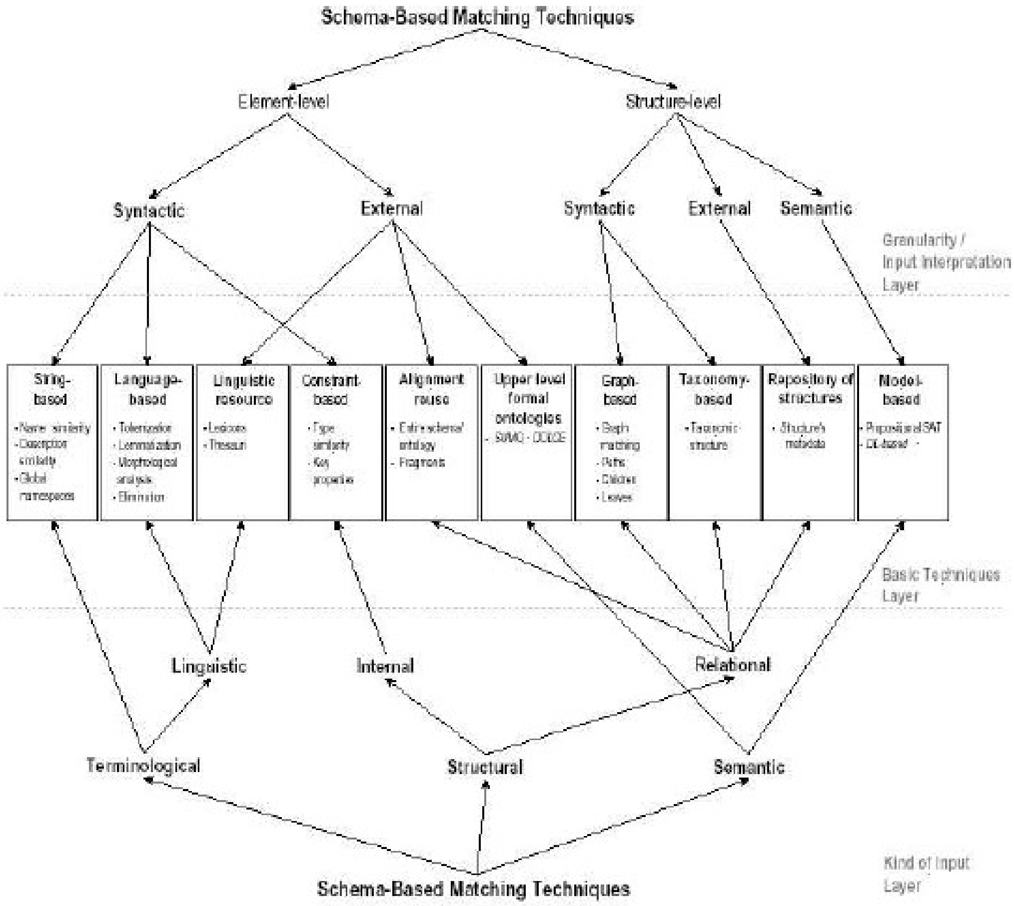


Figure 3. Classification of Giunchiglia and Shvaiko.

2.5 Proposed Classification

In our classification (Figure 4), as (Rahm & Bernstein, 2001), we distinguish between individual and combining matchers. The individual matcher can be on data level, ontology level, or context level.

At data level, the instances of the ontology are used as input to the matching approach.

At ontology level, the terms of the ontology structure and the hierarchy are taking into account. Then, we distinguish between element-level matcher and structure level matcher.

Moreover, the ontology's application context can be considered, i.e, how the ontology entities are used in some external context. This is specially interesting, for instance, to identify WordNet sense that must be considered to specific terms.

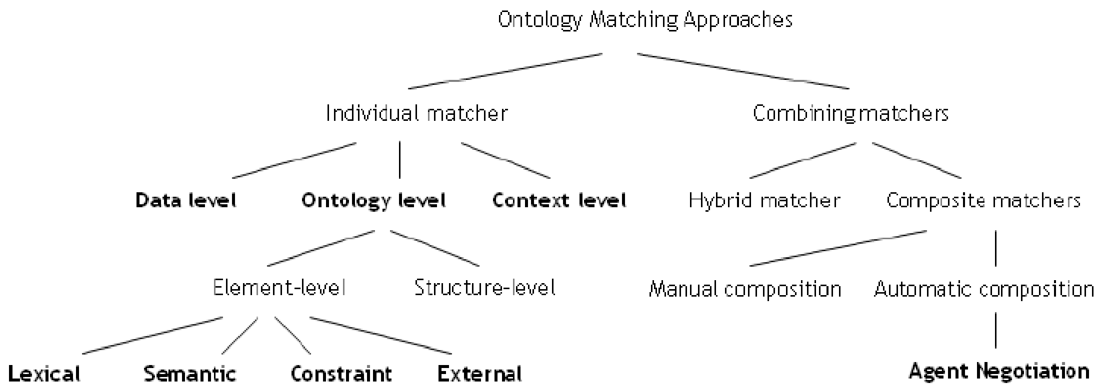


Figure 4. Proposed classification.

At element-level we consider lexical, semantic, structural, constraint-based, and external matchers.

The lexical approaches use metrics to compare string similarity are adopted. One well-known measure is the Levenstein distance or edit distance (Levenstein, 1966), which is given by the minimum number of operations (insertion, deletion, or substitution of a single character) needed to transform one string into another. Based on Levenshtein measure, (Maedche & Staab, 2002) proposes a lexical similarity measure for strings, the String Matching (SM), that considers the number of changes that must be made to change one string into the other and weighs the number of these changes against the length of the shortest string of these two. Others common metrics are: the Smith-Waterman (Smith & Waterman, 1981), which additionally uses an alphabet mapping to costs; and the (Euzenat et al., 2004) which searches for the largest common substring.

The semantic matchers consider semantic relations between concepts to measure the similarity between them, usually on the basis of one lexical database, thesaurus or similar semantic oriented linguistic resources. The well-known WordNet (http://www.wordnet.princeton.edu) database, a large repository of English lexical items, has been used to provide these relations. This kind of mapping is complementary to the pure string similarity metrics. It is not uncommon the cases where string metrics fail to identify high similarity between strings that represent completely different concepts. For example the words “score” and “store”, represent different concepts, but the Levenstein metrics return 0.68.

Constraint-based matchers are based on data types, value ranges, uniqueness, cardinalities, and others information constraint in the matching process. For example, the similarity between two terms can be based on the equivalence of data types and domains, of key characteristics (e.g., unique, primary, foreign), or relationship cardinality (e.g., 1:1 relationships) [19].

At element-level, we consider that external matchers consider some type of external information, such as user input or previous matching results.

The structural matchers use the ontology structure, where the positions of the terms in the ontology hierarchy are considered.

- Super(sub)-concepts rules. These matchers are based on rules capturing the above stated intuition. For example, if super-concepts are the same, the actual concepts are similar to each other. If sub-concepts are the same, the compared concepts are also similar (Dieng & Hug, 1998) (Ehrig & Sure, 2004).
- Bounded path matching (Noy & Musen, 2001): take two paths with links between classes defined by the hierarchical relations, compare terms and their positions along these paths, and identify similar terms. Example: Anchor-prompt. Anchor-prompt (Noy & Musen, 2001): takes as input a set of pairs of related terms anchors from the source ontologies. Either the user identifies the anchors manually or the system generates them automatically. From this set of previously identified anchors, Anchor-PROMPT produces a set of new pairs of semantically close terms. To do that, Anchor-PROMPT traverses the paths between the anchors in the corresponding ontologies. A path follows the links between classes defined by the hierarchical relations or by slots and their domains and ranges. Anchor-PROMPT then compares the terms along these paths to find similar terms.
- Leaves: the (structural) similarity between inner nodes of the graphs is computed based on similarity of leaf nodes, that is, two non-leaf schema elements are structurally similar if their leaf sets are highly similar, even if their immediate children are not, see, for example (Do & Rahm, 2001) (Madhavan et al., 2001). Example: TreeMatch. TreeMatch (Madhavan et al., 2001). (a) Atomic elements (leaves) in the two trees are similar if they are individually (linguistic and data type) similar, and if elements in their respective vicinities (ancestors and siblings) are similar. (b) Two non-leaf elements are similar if they are linguistically similar, and the subtrees rooted at the two elements are similar. (c) Two non-leaf schema elements are structurally similar if their leaf sets are highly similar, even if their immediate children are not. This is because the leaves represent the atomic data that the schema ultimately describes.

We also consider, as [19], hybrid and composite matchers, at combining matcher level. Hybrid matchers use multiple matching criteria (e.g., name and type equality) within an integrated matcher; and composite matchers (which can use a manual or automatic process) combine multiple match results produced by di®erent match algorithms. Our approach is an automatic composite matcher and then we add an agent negotiation approach at automatic level.

3. The Negotiation Model for Ontology Mapping

Negotiation is a process by which two or more parties make a joint decision (Zhang et al., 2005). It is a key form of interaction that enables groups of agents to arrive to mutual agreement regarding some belief, goal or plan (Beer et al., 1998). Hence the basic idea behind negotiation is reaching a consensus (Green et al., 1997).

A negotiation model is composed by two basic components (Lomuscio et al., 2001): the negotiation protocol and the negotiation strategies. The protocol specifies the rule of encounter between the negotiation participants. That is, it defines the circumstances under which the interaction between agents takes place, what deals can be made and what sequences of actions are allowed (Fatima et al., 2004). A strategy is the specification of the sequence of actions the agent plans to make during the negotiation.

Negotiation usually proceeds in a series of rounds, with every agent making a proposal at each round (Woo et al., 2002). The process can be described as follow (Lander et al., 1993). One agent generates a proposal and other agents review it. If some other agent does not like the proposal, it rejects the proposal and might generate a counter-proposal. If so, the others agents (including the agent that generated the first proposal) then review the counter-proposal and the process repeats. It is assumed that a proposal becomes a solution when it is accepted by all agents.

Cooperative negotiation is a particular kind of negotiation where agents cooperate and collaborate to obtain a common objective. It is a form of interaction where individual agents are cooperative and collaborate in order to achieve a common goal and for the best interests of the system as a whole (Beer et al., 1998) (Green et al., 1997). In cooperative negotiation, each agent has a partial view of the problem and the results are put together via negotiation trying to solve the conflicts posed by having only partial views (Gatti et al., 2004). According to (Guttman, 1998), cooperative negotiations can be described as the decision-making process of resolving a conflict involving two or more parties over multiple independent, but non-mutually exclusive goals.

This kind of negotiation has been currently adopted in resource and task allocation fields (Bigham et al., 2003) (Mailler et al., 2003) (Zhang et al., 2005). In these approaches, the agents try to reach the maximum global utility that takes into account the worth of all their activities. Differently from what is found in literature, in our approach the cooperative negotiation is a way to agents negotiate on a final mapping that is the result of different ontology mapping approaches.

In our model, the agents use lexical, semantic and structural approaches to map terms of two different ontologies. The distinct mapping results are shared, compared, chosen and agreed, and a final mapping result is obtained. This approach aims to overcome the drawbacks of the single ontology mapping approaches. First, we present the organization of the agents society and next we detail the negotiation process.

3.1 Organization of the Agents Society

We describe our model according to the agent society organization presented in Figure 5. The specification of the society organization is presented using the Moise+ model (Hubner, 2002). This model proposes three dimensions for organizations of agents society: structural, functional and deontic. The structural dimension defines the agents society structure describing what agents could do in their environments (their roles). The functional dimension defines the system functioning describing how agents execute their goals. The deontic dimension defines a kind of relation between structural and functional dimensions, describing the permissions and obligations of a role in a goal. This paper focused on the first dimension, presenting the structure specification for the organization of the proposed agents society.

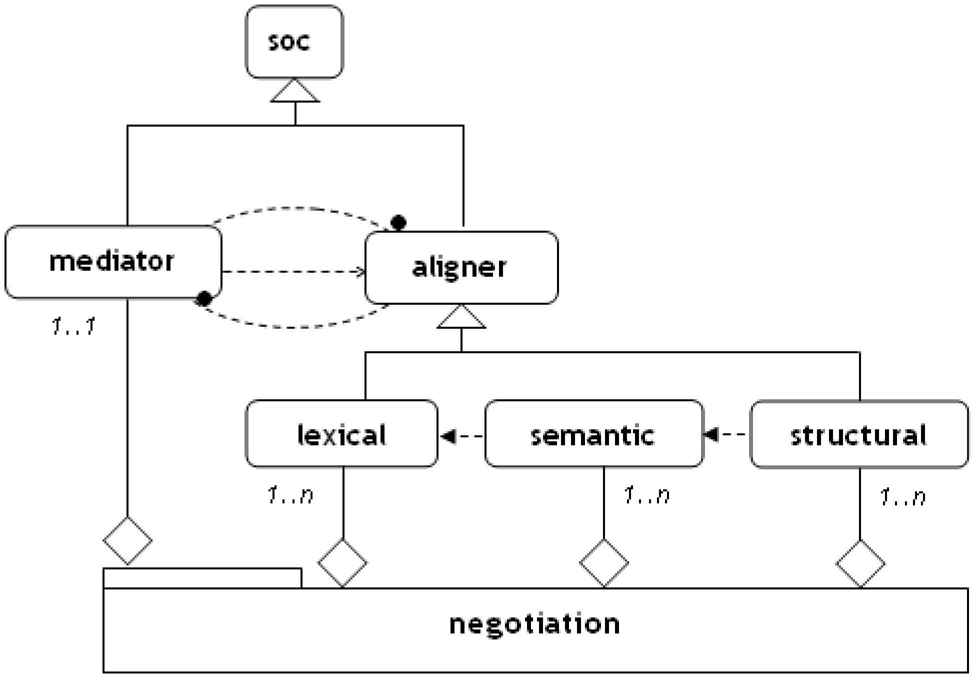


Figure 5: Organizational model.

According to (Hubner, 2002) and (Hubner, 2003), structural specification has three main concepts, roles, role relations and groups that are used to build, respectively, the individual, social and collective structural levels of an organization. The individual level is composed by the roles of the organization. A role means a set of constraints that an agent ought to follow when it accepts to play that role in a group. The following roles are identified in the proposed organization:

- Mediator: this role is responsible for mediating the negotiation process, sending and receiving messages to and from the mapping agents.

- Aligner: this role is responsible for giving an output between two ontology mappings (i.e., encapsule the mapping algorithms). One aligner could assume the role lexical, semantic or structural. On the lexical role, the aligner makes the mapping using algorithms based on string similarity. On the semantic role, the agent search by corresponding terms in a semantic oriented linguistic database. On the structural role, the agent is based on the intuition that if super-classes are the same, the compared classes are similar to each other. If sub-classes are the same, the compared classes are also similar.

In the social level the kinds of relations among roles that directly constrain the agents are defined. Some of the possible relations are:

- Acquaintance (acq): agents playing a source role are allowed to have a representation of the agents playing the destination role. In Figure 5, this kind of relation is present between the source role mediator and the destination role aligner and is represented by the graphical link, as indicated in Figure 6.

- Communication (com): agents playing a source role are allowed to communicate with agents that playing the destination role. In Figure 5 this kind of relation is present between the source role mediator and the destination role aligner (by heritage, lexical, semantic and structural) and is represented by the graphical link (Figure 6).
- Authority (aut): agent playing a source role has authority upon agent playing destination role. In Figure 5 this kind of relation is present between the source role semantic and the destination roles lexical and structural is represented by the graphical link (Figure 6).

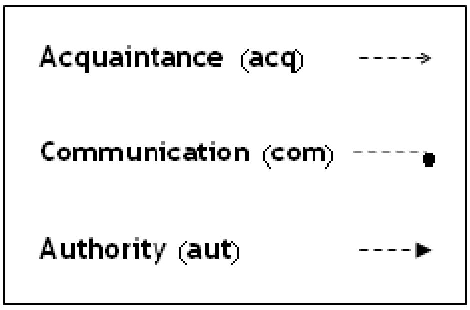


Figure 6. Relations among roles.

The collective level specifies the group formation inside the organization. A group is composed by the roles that the system could assume, the sub-groups that could be created inside a group, the links (relations) valid for agent and by the cardinality. A group can have intra-groups links and inter-groups links. The intra-group links state that an agent playing the link source role in a group is linked to all agents playing the destination role in the same group or in its sub-groups. The inter-group links state that an agent playing the source role is linked to all agents playing the destination role despite the groups these agents belongs to (Hubner, 2002). Links intra-group are represented by a hatched line and links inter-groups are represented by a continue line. This specification defines only a group called negotiation and all links are intra-group.

Based on the structural specification of the proposed organization, our society is composed by one agent that assumes the mediator role and three agents that assume the aligner role. One of the aligner agents is assuming the lexical role, one is assuming the semantic role, and one is assuming the structural role.

3.2 Negotiation Process

Basically, the negotiation process involves two phases. First, the agents work in an independent manner, applying a specific mapping approach and generating a set of negotiation objects. A negotiation object is a triple O = (T1,T2,C), where T1 corresponds to a term in the ontology 1, T2 corresponds to a term in the ontology 2, and C is the mapping category resulting from the mapping for these two terms.

Second, the set of negotiation objects, that compose the mapping is negotiated among the agents. The negotiation process involves one mediator and several aligner agents. In order to facilitate the negotiation process (i.e, reduce the number of negotiation rules), we define four mapping categories according to the output of the aligner agents.

Table 1 shows the categories and the corresponding mapping results. The output of lexical agents is a value from the interval [0,1], where 1 indicates high similarity between two terms (i.e, the strings are identical). This way, if the output is 1, “an mapping with certainty” is obtained. If the output is 0, the agent has a “not mapping with certainty”. A threshold is used to classify the output in uncertain categories. The threshold value can be specified by the user.

The semantic agents consider semantic relations between terms. Considering the WordNet database, relations such as synonym, antonym, holonym, meronym, hyponym, and hypernym can be returned. According to the mapping categories, synonymous terms are aligned with certainty; terms related by holonym, meronym, hyponym, or hypernym are considered aligned, but with uncertainty; when the terms can not be related by the WordNet (the terms are unknown for the WordNet database), the terms are not aligned, but with uncertainty.

The structural agent uses the super-classes intuition to verify if the terms can be considered similar. First, it is verified if the super-classes are lexically similar. Otherwise, the semantic similarity is used. If the super-classes are lexically or semantically similar, the terms are similar to each other. The matching category corresponds the output of the lexical or semantic comparison (e.g, if super-classes are not lexically similar, but they are considered synonymous, a “mapping with certainty” is returned).

Category	Lexical	Semantic
Mapping (certainty)	1	synonym
Mapping (uncertainty)	1 > r > t	related
Not mapping (uncertainty	0 < r <= t	unknown
Not mapping (certainty)	0	

Table 1. Matching categories.

Figure 7 shows an AUML interaction diagram with the messages changed between the agents during a negotiation round. We use an extension of AUML-2 standard to represent agents' actions (the actions are placed centered over the lifeline of the named agent). The interaction diagram refers to negotiation of the mapping between the classes "personal computer" and "pc"(Figure 8 and Figure 9)[1].

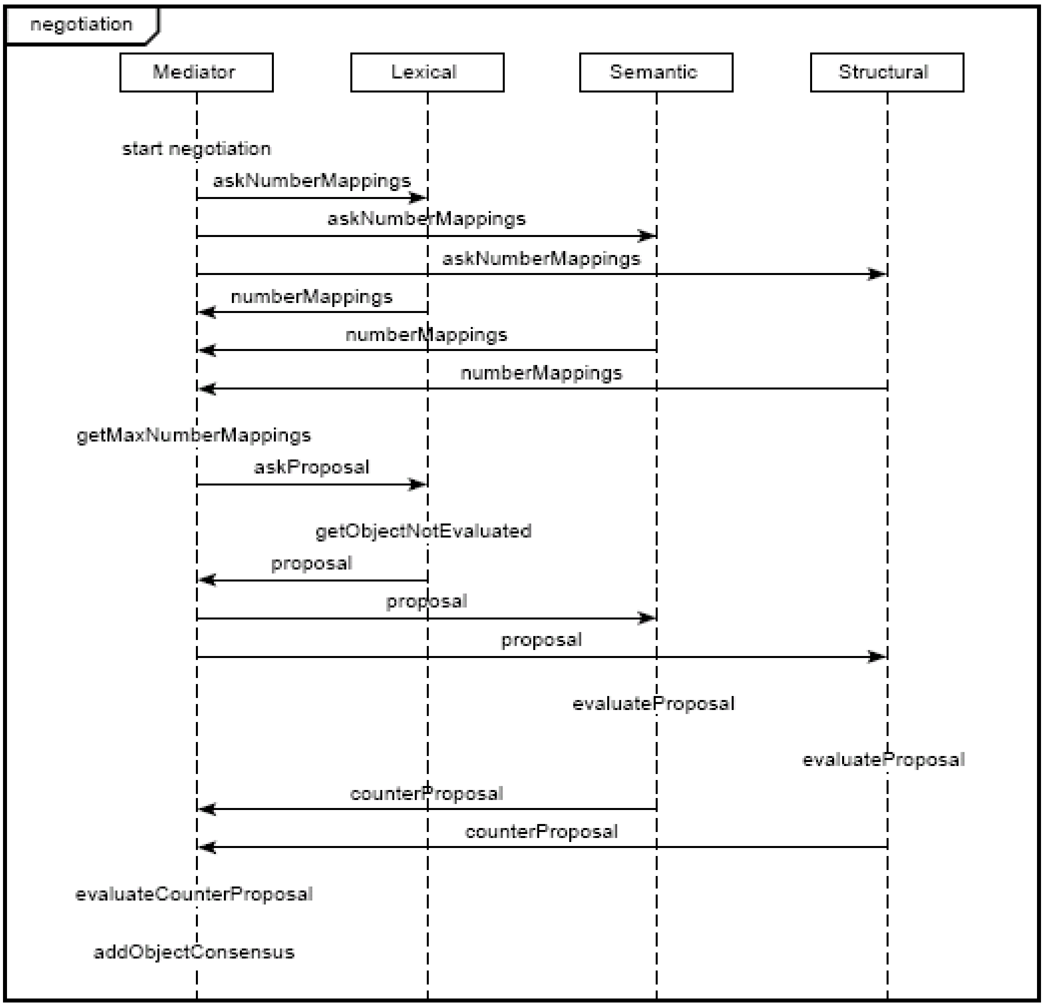
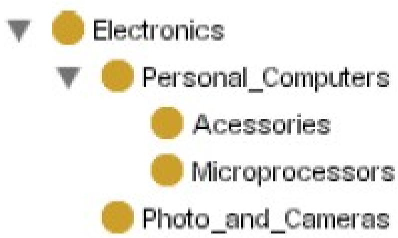


Figure 7. AUML negotiation interaction.



[1] Ontologies available in: http://dit.unitn.it/~accord/Experimentaldesign.html (Test 4)

Figure 8. Ontology 1.

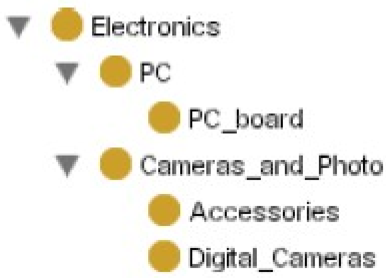


Figure 9. Ontology 2.

The negotiation process starts with the mediator agent asking to the aligner agents for its number of “mappings with certainty”. The first aligner agent to generate a proposal is one that has the greatest number of “mappings with certainty” (lexical agent, in the specific example).

The proposal contains the first negotiation object that still wasn’t evaluated by the agent. This proposal is then sent to the mediator agent, which sends it to others agents (in the specific example, the lexical agent proposes a “not mapping with certainty” to the mapping between the classes “personal computer” and “pc”). Each agent then evaluates the proposal, searching for an equivalent negotiation object. One negotiation object is equivalent to another when both refer to equals terms which are being compared in the two ontologies.

If an equivalent negotiation object has the same category, the agent accepts the proposal. Otherwise, if the agent has a different category for the compared terms in the negotiation object, its object negotiation is sent as a counter-proposal to the mediator agent, which evaluates the several counter-proposals received (several agents can send a counter-proposal). In the example, semantic and structural agents have generated counter-proposals, indicating a “mapping with certainty” between the compared terms. The semantic agent identifies that the terms are synonymous in WordNet, and structural agent identifies terms having the same super-class.

The mediator selects one counter-proposal that has the greater number of vote. If two categories receive the same number of votes, the category indicated by the semantic agent is considered a consensus.

The process is repeated. When a proposal is accepted by all agents or a counter-proposal consensus is obtained (through voting), the mediator adds the corresponding negotiation object in a consensus negotiation set and the aligner agents mark its equivalent one as evaluated. The negotiation ends when all negotiation objects were evaluated.

4. Final Remarks

Ontology mapping is an important problem to be dealt with in the context of the semantic web. It is also an important step in the process of ontology merging. It is a complex problem for which many different approaches have been developed. In this paper we present an agent oriented model for the composition of mapping approaches. We situate our model in a revised classification of ontology mapping, our model automatically combines approaches of different levels through a negotiation process. Given the nature of the problem we believe that this kind of composition is crucial for the establishment of a global ontology mapping, as it requires different knowledge levels, examples are lexical, semantic and structural knowledge. The advantages of the model is its adaptability to new approaches and easy of expansion. In future work we intend to consider argumentation for the enrichment of the negotiation process.

Acknowledgments. We would like to thank the financial support of CAPES/GRICES and Programme Alban, the European Union Programme of High Level Scholarships for Latin America (scholarship no. *E05D059374BR)*.

References

BEER, M.; D'INVERNO, M.; LUCK, M.; JENNINGS, N.; PREIST, C. and SCHROEDER, M. 1998. Negotiation in multi-agent systems. In *Workshop of the UK Special Interest Group on Multi-Agent Systems*.

BIGHAM, J. and DU, L. 2003. Cooperative negotiation in a multi-agent system for real- time load balancing of a mobile cellular network. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 568-575. ACM Press.

CHAVES, M. 2002. *Mapeamento e comparação de similaridade entre estruturas ontológicas*. Master’s Thesis, Pontifícia Universidade Católica do Rio Grande do Sul.

DIENG, R. and HUG, S. 1998. Comparison of personal ontologies represented through conceptual graphs. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 341-345.

DO, H-H. and RAHM, E. 2002. Coma - a system for flexible combination of schema matching approaches. In *Proceedings of the Very Large Data Bases Conference (VLDB)*, pages 610-621.

EHRIG, M. and SURE, Y. 2004. Ontology mapping - an integrated approach. In *Proceedings of the European Semantic Web Symposium*, pages 76-91.

EUZENAT, J.; LE BACH, T.; BARRASA, J.; BOUQUET, P. ; DE BO, J.; DIENG-KUNTZ, R.; EHRIG, M.; HAUSWIRTH, M.; JARRAR, M.; LARA, R.; MAYNARD, D.; NAPOLI, A.; STAMOU, G.; STUCKENSCHMIDT, H.; SHVAIKO, P.; TESSARIS, S.; VAN ACKER, S. and ZAIHRAYEU, I. *State of the art on ontology alignment*. Deliverable 2.2.3, 2004.

FATIMA, S.; WOOLDRIDGE, M. and JENNINGS ,N. 2004. An agenda-based framework for multi-issue negotiation. *Artificial Intelligence*, 152:1-45.

GATTI, N. and AMIGONI, F. 2004. A cooperative negotiation protocol for physiological model combination. In *Proceedings of the Third Internation Joint Conference on Automomous Agents and Multi-Agent Systems*, pages 655-662.

GIUNCHIGLIA, F. and SHVAIKO, P. 2004. Semantic matching. *Knowledge Engineering Review*, 18(3):265-280.

GREEN, S.; HURST, L.; NANGLE, B.; CUNNINGHAM, P.; SOMERS, F. and EVANS, R. 1997. *Software agents: A review*. Technical report, Trinity College.

GUTTMAN, R. H. and MAES, P. 1998. Cooperative vs. competitive multi-agent negotiations in retail electronic commerce. *Lecture Notes in Computer Science*, 1435:135-147.

HOVY, E. 1998. Combining and standardizing large-scale, practical ontologies for machine translation and other uses. In *Proceedings of the First International conference on language resources and evaluation.*

HUBNER, J. 2003. *Um Modelo de Reorganização de Sistemas Multiagentes*. PhD thesis, Escola Politécnica da Universidades de São Paulo, Departamento de Engenharia da Computação e Sistemas Digitais.

HUBNER, J.; SICHMAN, J. and BOISSER, O. 2002. A model for structural, functional, and deontic specification of organizations in multiagent systems. *In Advances in Artificial Intelligence: 16th Brazilian Symposium on Artificial Intelligence, SBIA - Lecture Notes in Computer Science 2507, Springer.*

LEVENSTEIN, L. I. 1966. *Binary codes capable of correcting deletions, insertions and reversals*. In Cybernetics and Control Theory.

MADHAVAN, P. B. J. and RAHM, E. 2001. Generic schema matching with cupid. In *Proceedings of the Very Large Data Bases Conference (VLDB)*, pages 49-58, 2001.

LANDER, S. and LESSER, V. 1993. Understanding the role of negotiation in distributed search among heterogeneous agents. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

LOMUSCIO, A. R., WOOLDRIDGE, M. and JENNINGS, N. 2001. A classification scheme for negotiation in electronic commerce. *Agent-Mediated Electronic Commerce: A European Agent Link Perspective*, pages 19-33.

MAEDCHE, A. and STAAB, S. 2002. Measuring similarity between ontologies. In *Proceedings of the European Conference on Knowledge Acquisition and Management*, pages 251-263.

MAILLER, M.; LESSER, V. and HORLING, B. 2003. Cooperative negotiation for soft real-time distributed resource allocation. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 576-583. ACM Press.

NOY, N. and MUSEN, M. 2001. Anchor-prompt: using non-local context for semantic matching. In *Proceedings of the workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence* (IJCAI), pages 63-70.

RAHM, E. and BERNSTEIN, P. A. 2001. A survey of approaches to automatic schema matching. *VLDB*, 10:334-350.

SHVAIKO, P. 2004. *A classification of schema-based matching approaches*. Technical report, Informatica e Telecomunicazioni, University of Trento.

SHVAIKO, P. and EUZENAT, J. 2004. *A survey of schema-based matching approaches*. Technical report, Informatica e Telecomunicazioni, University of Trento.

SMITH, T. and Waterman, M. 1981. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147:195-197.

WOOLDRIDGE, M. 2002. *An Introduction to Multiagent Systems*. John Wiley and Sons.

ZHANG, X.; LESSER, V. and PODOROZHNY, R. 2005. Multi-dimensional, multistep negoriation for task allocation in a cooperative system. *Autonomous Agents and Multi-Agent Systems*, 10:5-40.