

A socio-semantic approach to collaborative domain conceptualization

Carla Pereira^{1,2}, Cristóvão Sousa^{1,2}, António Lucas Soares^{1,3}

¹Instituto de Engenharia de Sistemas e Computadores do Porto (INESC Porto),
Campus da FEUP, Rua Dr. Roberto Frias, 378, 4200 - 465 Porto, Portugal

²Escola Superior de Tecnologia e Gestão de Felgueiras – Instituto Politécnico do Porto,
Rua do Curral, Casa do Curral, Margaride, 4610-156, Felgueiras, Portugal

³Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465,
Porto, Portugal

csp@inescporto.pt, cpsousa@inescporto.pt, als@fe.up.pt

Abstract. Current knowledge about the early phases of ontology construction is insufficient to support methods and techniques for a collaborative construction of a conceptualization. Following an approach inspired in cognitive semantics, the application and extension of the Conceptual Blending Theory (CBT) to the realm of collaborative semantic tools is proposed. A formal framework is presented together with some examples in the collaborative networks context. A collaborative semantic architecture is presented. This architecture supports the two components of the proposed method: collective conceptualization and consensus reaching.

Keywords: Conceptualization, social construction of meaning, ontology development, collaborative networks, conceptual blending.

1 Introduction

Even though the most used definition of ontology [4] "An ontology is a formal, explicit specification of a shared conceptualization", underlines the collaborative construction of conceptualizations in the scientific context, we agree that: "While different degrees of formalization have been well investigated and are now found in various ontology-based technologies, the notion of a shared conceptualization is neither well-explored, nor well-understood, nor well-supported by most ontology engineering tools" [8]. Current knowledge about the early phases of ontology construction is insufficient to support methods and techniques for a collaborative construction of a conceptualization [5]. Nevertheless, the conceptualization phase is of utmost importance for the success of the ontology, as it is in this phase that a socio-semantic agreement is shaped. Our view is that ontology engineering needs a "socio-cognitive turn" in order to generate tools that are really effective in copying the complex, unstructured, and highly situational contexts that characterize a great deal of information and knowledge sharing in business collaboration. As it argued in [1]: "we need to go beyond the approaches that provide a high level of 'automation of the

meaning'; instead, we need to address situations where human beings are highly required to stay in the process, interacting during the whole life-cycle of applications, for cognitive and cooperative reasons".

The research work described in this paper was motivated by the need to support effectively information and knowledge sharing in the context of collaborative networks with short life-cycles. The early phases of setting up information and knowledge management architectures for inter-organizational teams (e.g., large scale international R&D or systems engineering projects) are complex and problematic due, in a great extent, to the actor's heterogeneous professional and cultural backgrounds. A good example of this is the construction industry, where an enormous effort and money has been spent in standard terminologies, vocabularies, thesaurus, ontologies, etc. with results well behind the expected [7].

This line of research is therefore directed towards the application of cognitive semantic results in the creation of artifacts acting as socio-technical devices supporting the view that meaning is socially constructed through collaboration and negotiation. The first line of this research work deals with the application and extension of the Conceptual Blending Theory (CBT) [3] to the realm of collaborative semantic tools. The practical application of our approach is to support the co-construction of semantic artifacts by groups of social actors placed in organizational contexts interacting towards a set of common objectives. Simple examples of these artifacts are the creation of a common taxonomy (or ontology) for classifying and retrieving content from an inter-organizational portal, the creation of specific terminological accounts to serve as conceptual references in project tasks, or the specification of ontologies for systems interoperability.

2 The Conceptual Blending Theory

Our proposal to support a collaborative process of conceptualization of a given reality (e.g., a domain in the context of a project) is founded on cognitive semantics, specifically on the Conceptual Blending Theory [3]. CBT accounts for the emergence of meanings by adopting the view that meaning construction involves emergent structure, i.e., conceptual integration is more than the sum of its component parts. An integration network is thus a mechanism for modeling how emergent meaning might come about, accounting for the dynamic aspects of meaning construction.

CBT representation gives rise to complex networks by linking two (or more) input spaces by means of a generic space. The generic space provides information that is abstract enough to be common to all the input spaces. Elements in the generic space are mapped onto counterparts in each of the input spaces, which motivate the identification of cross-space counterparts in the input spaces. A further space in this model of integration network is the blended space or blend. This is the space that contains new or emergent structure. The blend takes elements from both inputs, but goes further on providing additional structure that distinguishes the blend from either of its inputs. In other words, the blend derives a structure that is contained in neither input (see [2] and [5]).

3 A method to support the conceptualization process

The several case studies accomplished in the scope of our participation in two European projects, AC+DC¹ and H-Know², allow us to refine and improve our method as well as to propose a tool to support it (case studies described in [6] and [10]).

3.1 Collaborative network characterization

Departing from the work of Roth [9], we develop a conceptual and formal framework for modeling collaborative networks as epistemic networks. Our view is that an epistemic network is made of three networks: a social network (involving links between individuals), a semantic network (with links between concepts) and a socio-semantic network (which links individuals to concepts). At the end all partners that belong to the network will be linked to concepts in the same way, or rather, to the semantic network that represents the shared conceptualization. In our context, the nodes in a social network represent the set of individuals of the several organizations (partners), that compose the collaborative network, and links represent the relationships between the several individuals. The semantic network is the network of concepts that represents the strategic frame of the collaborative network, where nodes are concepts of the several domains that compose the network and edges are relations between concepts. The final semantic network will be the shared conceptualization presented in the generic space (considering the network model of CBT theory [5]). The socio-semantic network is made of individuals of the several organizations that compose the collaborative network, concepts of the several business domains, and links between them, representing the usage of concepts by individuals within the strategic frame. The final socio-semantic network represents the relationship between all organizations that compose the network and the shared conceptualization. Considering the three networks, we deal with different kinds of links that influence the result of the shared conceptualization process. Considering the social network, we consider two kinds of relations: links between the several organizations that constitute the collaborative network, i.e., between individuals from different organizations; (these links influence during the negotiation process); and links between pairs of individuals within the organization (these links influence the input spaces creation. However, at this moment, this relationship will not be addressed by us). Considering the semantic network, we deal with two kinds of relations: links between concepts of different conceptualization proposals (links between input spaces – considered in the step 2 of our method, see next section) and links between concepts within each conceptualization proposal (each input space). Considering the socio-semantic network, we deal with the links between the individuals that compose the network and the concepts that compose the shared conceptualization.

¹ <http://www.acdc-project.org/public/>

² <http://dionisio.inescporto.pt:8282/hknow>

3.2 CBT based method

Steps. With a set of assumptions already described in [5], the proposed method establishes the following steps (see figure 1). *Step 0:* Based on the definition of the strategic frame a subgroup of participants should create a preliminary proposal that will be shared in the generic space and used as foundation for the step 1. *Step 1:* each organization has assigned one or more input spaces (only one input space per organization is considered here, for simplicity) and represents its conceptualization proposal through the input space. Considering the definition of “space”, each conceptualization proposal is a concept map with the information and other knowledge sources which allow the correct understanding of the conceptualization proposal. *Step 2:* a generic conceptualization is generated; the common conceptual structure in the generic space should be broad enough to be accepted by all the team members with minimum negotiation. *Step 3:* considering the “counterpart” elements, the process of creating the blend space is started using selective projection; based on the input spaces, strategic frame, generic space and documentation available in the input spaces (called background information), the blend is “run” to obtain new conceptualization proposals. *Step 4:* new conceptual structures proposed in the blend space are object of negotiation; the concepts for which consensus exists are represented in (“copied” to) the generic space; situations that justify “backward projection” to the input spaces and their modification are analyzed (this analysis will be performed by the users, after obtaining consensus) then the emergent blend structure is validated (confirm or eliminate new concepts that raise in the blend). *Step 5:* if input spaces modification takes place, the method should resume at step 2; however the creation of a new blend space, is not necessary.

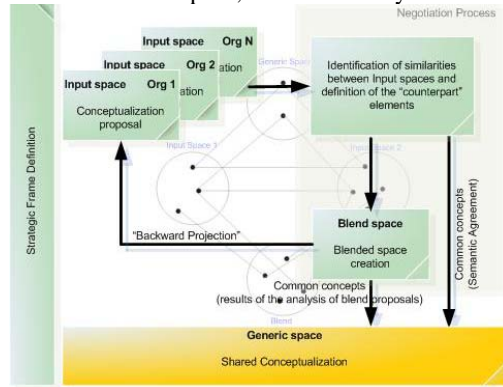


Fig.1. Method to support the collaborative conceptualization process.

Step 6: when all participants manifest their agreement with the conceptualization represented in the generic space, the method instance is finished. Whenever an organization introduces a new concept or association in their input space, it becomes a target for negotiation. At this moment the negotiation process is started. The network team evaluates the new entrance, and if necessary “run” the steps 2, 3 and 4 to help their decision.

Consensus Building Strategy to support the Negotiation Process.

To ensure a successful negotiation process, which lead to a shared conceptualization (accepted by all partners), is also very important to define an approach for obtaining consensus. For this goal, there are some good practices that should be applied, but that need to be adjusted for each case. In [10] we present a strategy for building consensus in the context of collaborative networks, aligned with the fact that the process of building consensus is regarded collaborative right from the design, once the process design itself should involve all participants [11]. Our approach to reach a semantic agreement for the collective conceptualization is thus based on consensus building techniques and involves, besides the negotiation itself, a preparation and an evaluation phases (see [10]). For the preparation phase, the creation of the team and the governance model of the collaboration are the main tasks. In the consensus building phase, must be used a combination of the following two approaches to drive the process, “working with a single-text document” and “taking a visioning approach”. The single-text approach involves introducing a working draft of an agreement early in the process for parties to discuss and revise. This approach provides a clear structure for discussions and a focal point to identify areas of agreement and disagreement. A subgroup of participants works to draft a preliminary proposal. This preliminary proposal is considered the best way to focus a consensus building dialogue (performed in the step 0 of the method). Brainstorming will be used to expand and improve the preliminary proposal using the proposals presented in the input spaces of each organization and to encourage creative thinking. The exploration of several proposals should be accomplished using questions of the type “what if” and performing the steps 2, 3 and 4 of the method to support the collaborative conceptualization process. The “taking a visioning approach” intend to focus participants’ attention toward the future in the course of identifying options and seeking agreements. Conduct the process according to the following questions: What do you have?, What do we want? And “How do we get there? These questions allow the semantic team evaluates the strengths and weaknesses of the current conceptualization proposal and to maintain the process more creative. To support the dissemination of information, a content management system and/or wiki system should be used, that allows notifying the participants in the process, to distribute materials to participants during the process, allow the information sharing, keep all stakeholders updated about the group progress and keep a record of all items discussed. A set of documents such as: meeting minutes, agendas, schedules, working documents and background resources, are taken as important to share. If necessary, at this stage, the agenda and procedural rules can be modified. Evaluation and lessons learnt, before closing the process, the final version of the conceptualization should circulate for all participants. At this moment should be organized a workshop to divulge the final result to all partners. The evaluation of strengths and weaknesses of the process will be performed based on experience obtained through our participation in the process and carrying out an online questionnaire to all participants. Responsibilities and roles definition: There are three different roles that the participants can assume (see figure 2): partner (any member of the collaborative network may take this role); reviewer (each organization define a representative who will assume the role of reviewer; the representatives are part of the group responsible for making decisions). When the specificity of the conceptualization increases the

need to create different work groups will be evaluated again; and mediator, a person responsible for ensuring the proper participation of all and managing the conceptualization. The consensus building workflow used is presented in figure 3 and described in [10].

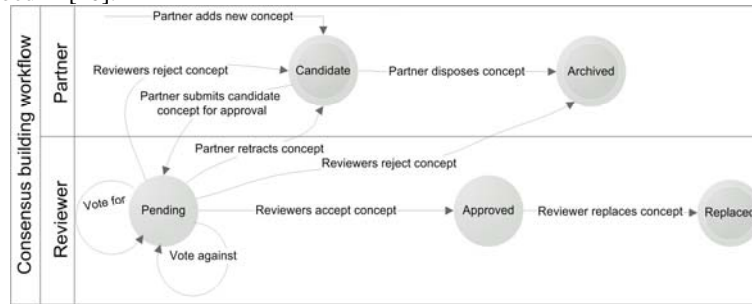


Fig.2. Consensus building workflow

Techniques and Tools. During our research work and based on the case studies carried out through an action research approach, some tools and techniques were placed in evidence to support the proposed method. We are now able to summarize those techniques and supporting tools as well as the current architecture which supports our method. From the terminological domain we used the lexical database, IATE³, the inter-institutional terminology database of the European Union and TermExtractor⁴. IATE allowed us to search for the correct correspondent terms chosen to name a concept, in several languages, according with a specific domain, along the concept maps creation within each input space. TermExtractor, is a text mining tool, used to extract relevant terms in the interest domain, by submitting an archive of domain-related documents in any format. Those documents used were mainly internal project documents, the logistics area of SAP dictionary⁵, the terms and glossary about supply chain management proposed by [12]. As the knowledge representation technique for joint construction of a conceptual representation is proposed the use of concept maps, supported by CmapTools⁶ software. The use of concept maps would allow untrained users, that cannot be expected to conform to the constraints of the formal semantics, to concentrate on the task at hand in an unrestricted environment. Our goal is to allow users (all participants in the conceptualization process) could start informally, without having to translate their know-how into any knowledge representation (heavy) formalism [13]. The exploration of CmapTools features, lead us to clarify some socio-technical assumptions, as well as specify and/or refine new ones in order to achieve an architecture able to support the proposed CBT methodology. CmapTools was

³ <http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load>

⁴ <http://lcl2.uniroma1.it/termextractor/>

⁵

http://help.sap.com/saphelp_46c/helpdata/En/35/2cd77bd7705394e10000009b387c12/frameset.htm

⁶ <http://CMap.ihmc.us/conceptmap.html>

essential to support the frame definition and input and blend services were supported by Plone CMS and Semantic Media Wiki (SMW). CmapTools can actually support, partially, the proposed method (frame definition and input spaces), however we are facing an inter-organizational processes with several steps and stages, which asks for a more comprehensive and collaborative environment architecture (see Fig. 3.), calling for several specific services in order to support consensus building and negotiation processes.

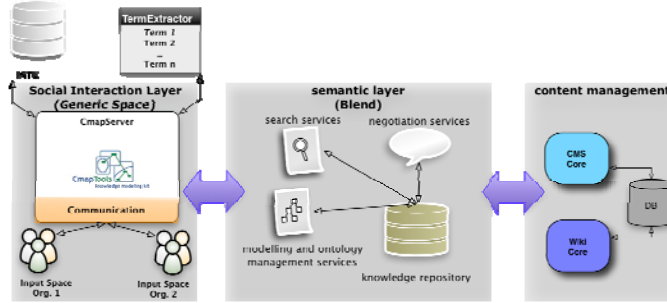


Fig.3. Dynamic Collaborative Environment Architecture

Following this, and according to the goals described earlier on this paper, (where it is mentioned the need to disseminate the information, towards to allow the semantic team to evaluates the strengths and weaknesses of the current conceptualization proposal) two modules are to be developed: one supporting the semi-automated construction of the conceptual integration spaces (semBlend) and the other supporting semantic negotiation and consensus reaching (semCons). These make part of a broader platform: SemSys4CCM (Semantic System for Continuous Construction of Meaning). Currently several services are running in order to communicate with CmapServer, and getting the conceptualization proposal developed according with the initial steps of the method. With this feature we are able to have all the input spaces in our knowledge repository. At this point, the functionalities of the semantic media wiki are used to discuss the input spaces and to annotate the concepts within each input space, gathering information to run the blend. Additional features were developed to allow, maintenance, navigation and integration and synchronization with the CmapServer.

4 A Formal Framework

Definition 1: Concept map. Let us consider LC as a set of symbols (the concepts) and LR as set of symbols (the relations). A concept map CM is a tuple (LC, LR, LCM), where LCM is a set of literals of the form $r(y, z)$, such that $r \in LR$; $y, z \in LC$.

$$CM = \{r(y, z) : r \in LR, y, z \in LC\} \quad (1)$$

Definition 2: Space. A conceptual space, E , is a tuple (CM, LK) , where CM is a concept map and LK is a set of links to information and knowledge sources.

Definition 3: Context Frame. Given a collaborative network, N , the context frame (F_N) can be defined as a tuple (D_N, O_N) , where D_N are the set of business domains of the organizations that compose the network, O_N are the set of goals of the network (common organization's goals together with specific goals for the network).

Definition 4: Conceptualization proposal (input space).

Let us consider $T \subseteq S \times D$ binding S to D . We introduce the operation “ \wedge ” such that for any element $s \in S$, s^\wedge is the set of elements of D_N which are T – related to s . T expresses any kind of relationship between an organization s and business domain d . In our case, this relationship represents the fact that d is one of the business domains of s .

$$s^\wedge = \{d \in D \mid sTd\}. \quad (2)$$

Let us consider $\beta \subseteq d \times F_N$ binding d to F_N . We introduce the operation “ $*$ ” such that for any element $c, k \in d : d \subset D$, x^* is the set of elements of d which are β – related to F_N . β expresses the fact that a subset of LC or a subset of LK d to be part of F_N . c represents a concept and k represents a information source of the domain d .

$$(x^*)^c = \{c \in d \mid c\beta F_N\} \quad (3)$$

$$(x^*)^k = \{k \in d \mid k\beta F_N\}$$

When we start the creation of a conceptualization proposal, we use the information and knowledge of each domain and organize them in a space. Being, the input space of an organization s_i , a tuple $(LC_1', LK_1' LR', LCM') = (CM_1, LK_1)$, where LC_1' is the set of concept, $(x^*)^c$, of the business domains of s_i related to F_N , LK_1' is the set of information sources, $(x^*)^k$, of the business domains of s_i related to F_N .

A first version of the shared conceptualization available for discussion (presented in the blend space), G' , created with focus on the several partners proposals can be formalized as follow. The concepts for which consensus exist are copied for the generic space. G' is created using the mapping operation, φ , that find a set of mappings between two concepts presented in different proposals (in different input spaces) and the projection operation η . For lack of space, the several operations will not be described with detail here. The final shared conceptualization, G , is a tuple $(LC_G, LR_G, LK_G, LCM_G)$ where LC_G and LR_G are composed by elements (concepts or relations, respectively) from the mapping operation, blend creation (the elements of b^Δ) and elements proposed during the negotiation process by the participants.

Definition 5: Negotiation and decision-making space (Blend space). The blend space (B) results from the application of the composition, completion and elaboration processes to the input spaces. It is a tuple (LC_B, LR_B, LCM_B) , where LC_B and LR_B results from the projection and completion operations. A projection operation determines, for each element of a concept map in the input spaces, what will be its correspondent in the blend space. The input for the projection operation is the result of the mapping function application. The completion operation provides new elements, not visible in the input spaces but that are part of F_N , with base in some filters applied to the information sources available in the several spaces. That is, this space provides a new proposal of shared conceptualization for negotiation. For us, a blend is defined by the projection (η) and completion (δ) operations. In an initial iteration, we consider the results of the projection operation as the first version of the shared conceptualization (G'). During the negotiation process the participants can “run the blend”, which corresponds to the completion operation. Considering a proposed blend space, B , the several organizations negotiate the proposal for the shared conceptualization. Like this, let us consider $\theta \subseteq LC_B \times S$ binding LC_B to S . We introduce the operation " Δ " such that for any element $s \in S$, b^Δ is the set of elements of LC_B θ -related to every element of S . In our case, the θ relationship represents the existence of consensus between all organizations S about a $c \in LC_B$.

$$b^\Delta = \{c \in LC_B \mid \forall s \in S, c\theta s\} . \quad (4)$$

During the negotiation process can be proposed others concepts for besides the ones that constitute the input spaces and blend space. Of these proposals, the concepts for which consensus exists are also presented in the generic space.

5 Conclusions

The approach described in this paper proposes a shift in the process of creation of semantic artifacts from a “semantic artifact engineering” perspective to a “actor-artifact co-evolution” one. Socio-semantics is the scientific umbrella to this approach, which is also inspired in cognitive semantics and networking social theories. A set of detailed case studies is under way, whereas action research was adopted as the privileged way to generate knowledge in socio-technical settings. The results already achieved through case studies allowed us to design the architecture of a collaborative environment architecture that will support all steps of the method.

Future work involves the refinement of the architecture. Further developments are needed to meet the requirements for a semi-automated conceptual integration. This will require the development of semantic artifacts over the content management layer, extending a wiki to a semantic level and providing a set of services which, in

conjunction with CmapTools, will offer a suitable environment to support the negotiation and the Blend space.

References

1. Cahier, J-P., Zaher, L'H., Leboeuf, J-P, Guittard, C.: Experimentation of a socially constructed "Topic Map" by the OSS community. Proceedings of the IJCAI-05 workshop on KMOM, Edimbourg, (2005).
2. Evans, V, Green, M.: Cognitive Linguistics: an introduction. Edinburgh University Press, (2006).
3. Fauconnier, G., Turner, M.: Conceptual Integration Networks. Published in Cognitive Science, vol. 22 (2), pp. 133 – 187, (1998).
4. Gruber, TR.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5 (2), pp. 199-221, (1993).
5. Pereira, C., Soares, A.: Ontology development in collaborative networks as a process of social construction of meaning. On the Move to Meaningful Internet Systems: OTM 2008 Workshops, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, November, (2008).
6. Pereira, C., Sousa, C., Soares, A.: A socio-semantic approach to the conceptualisation of domains, processes and tasks in large projects. 11th International Conference on Enterprise Information Systems, 6 - 10, May, (2009).
7. Silva, M., Soares, A., Simões, D.: Selecting and Structuring Semantic Resources to Support SME'S Knowledge Communities. INCOM 2006, 12th IFAC Symposium on Information Control Problems in Manufacturing. St. Etienne, France, (2006).
8. Staab, S.: On understanding the collaborative construction of conceptualisations. International and Interdisciplinary Conference "Processing Text-Technological Resources" at the Center for Interdisciplinary Research (ZiF - Zentrum für interdisziplinäre Forschung), Bielefeld University, 13-15 March, (2008).
9. Roth, C.: Co-evolution in Epistemic Networks Reconstructing Social Complex Systems. Structure and Dynamics: eJournal of Anthropological and Related Sciences, vol. 1 (3), (2006).
10. Pereira, C., Sousa, C., Soares, A.: Short-term semantic consensus: towards agile ontology specification for collaborative networks, 10th IFIP Working Conference on VIRTUAL ENTERPRISES (PRO-VE'09), Thessaloniki, GREECE, 7-9 October (2009).
11. Susskind, L, McKearnan, S, Thomas-Larmer, J.: The Consensus Building Handbook: A comprehensive Guide to Reaching Agreement. The Consensus Building Institute. Sage Publications, (1999).
12. Vitasek, K.: Supply Chain Management, Terms and Glossary, (2008).
13. Eskridge, T., Hayes, P., Hoffman, R., Warren, M.: Formalizing the informal: a confluence of concept mapping and the semantic web. Concept Maps: Theory, Methodology, Technology. Proc. of the Second Int. Conference on Concept Mapping, A. J. Cañas, J. D. Novak, Eds., Costa Rica (2006).
14. Pereira, F.: Creativity and Artificial Intelligence: A conceptual Blending Approach. Applications of Cognitive Linguistics, Mouton de Gruyter, Berlin/New York (2007).