On Leveraging Schema Matching Networks for Explorative Process Analysis

Avigdor Gal and Matthias Weidlich*

Technion – Israel Institute of Technology, Haifa, Israel
avigal@ie.technion.ac.il, weidlich@tx.technion.ac.il

Abstract. Business networks have a large potential of supporting their members in identifying synergies, building up consortia, and joining forces to react to legal changes. Hence, business networks should be equipped with an infrastructure supporting those goals. Bridging data heterogeneity and enabling interoperability of business documents in such a setting was recently addressed by schema matching networks. Matching in such a network is guided by reusable building blocks, called concepts, that are shared by many network members. We propose to adapt the notion of schema matching networks for an exploration of the business processes of network members. Then, a network consists of abstract process descriptions and matching is guided by the business objects to which the activities refer.

1 Introduction

Collaboration is a key to react to technological trends, market shifts, or changes of legal regulations. This holds for small and medium size enterprises (SMEs) in particular, since they often lack the possibility to assess the impact of a market change. Collaboration helps SMEs to mitigate the risk of reacting inappropriately, e.g., by benefiting from the experience of early adopters or building up consortia to lower individual investments.

Business associations emerged as organisational bodies to structure the collaboration of SMEs that have common interests and, therefore, form a business network. The question of how infrastructure can foster collaboration within a business network has recently been addressed on the level of data heterogeneity. Schema matching, the act of generating correspondences between attributes of data schemas, has been applied to business documents exchanged within a network.[1] The underlying model is a schema matching network that comprises data schemas and mappings between them, similar to peer data management systems [1]. It allows for exploring shared business documents and, eventually, supports interoperability. Matching within the network is guided by reusable building blocks, i.e., concepts [2] that are shared by many network members. By decomposing schemas into concepts, one obtains a common interpretation of schemas in the network, which allows for conclusions on collaboration opportunities.

* The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 256955.

[1] Schema matching to enable document interoperability is investigated in the NisB project, see http://www.nisb-project.eu/.

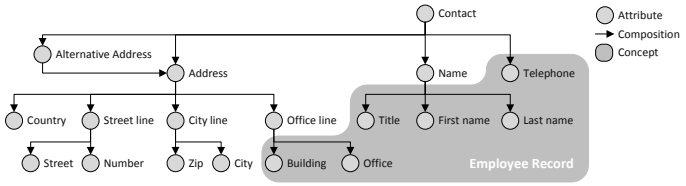

Fig. 1. A data schema represented as an attribute graph; a concept is highlighted.

In this paper, we argue that a network infrastructure for exploring a business network shall not be limited to the data level. Instead, lifting the ideas on schema matching networks to a business processes creates further opportunities for fostering collaboration. We outline a model, in which abstract descriptions of business processes are matched within a business network. Common knowledge shared within the network is built around business objects handled in these processes. Then, a *process matching network* allows for exploring operations implemented, envisioned, planned, or enquired by the members of a business network.

The next section summarises ideas on schema matching networks. Section 3 discusses process matching networks. Section 4 concludes and outlines steps to realize our vision.

2 Schema Matching Networks

Since decades, schema matching provides methods to cope with data heterogeneity by supporting the generation of correspondences between schema attributes [3]. Recently, it has been recognised as a tool to explore data commonalities and differences, e.g., for decision making [4]. In contrast to classical data integration, here, the assumption of validating all correspondences by a human expert no longer holds and a certain level of uncertainty is acceptable. Also, schema matching evolved from pairs of schemas in isolation to the network level. To cope with the challenges induced by a network, e.g., global match consistency, reusable building blocks, aka concepts [2], guide the matching task and provide a common interpretation of schemas within the network.

Schema matching networks emerged as a model to enable *exploration* of data schemas within a *network*. We illustrate this model with an example. Figure 1 depicts a data schema as an attribute graph. Nodes represent attributes and directed edges model composition of attribute types. An attribute graph is not necessarily tree-structured, e.g., a 'Contact' may comprise two address entries. Once an attribute graph is considered within a network, it is decomposed based on concepts that are already known in the network. Those concepts may stem from concept libraries, such as the Universal Business Language (UBL)[2]. For the example, a concept-based decomposition may identify that the attributes highlighted in Fig. 1 form a reusable block, i.e., an 'Employee record'.

Concepts are used to match a data schema with other schemas in a network. The obtained model is schematically illustrated in Fig. 2. Attribute graphs are connected by correspondences that are defined between attributes. A correspondence has a quality

[2] http://docs.oasis-open.org/ubl/UBL-2.1.pdf

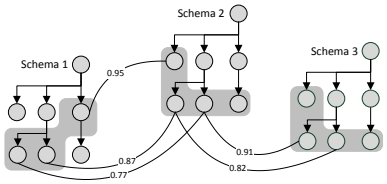

Fig. 2. Illustration of a schema matching network.

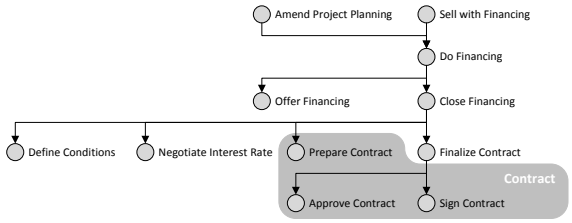

Fig. 3. An abstract representation of a financing process as an activity graph. It may be decomposed according to the artefacts that the activities refer to, such as the contract object.

value that represents its certainty. Different strategies may be followed for schema decomposition and concept-aware matching, involving, e.g., concept filtering and handling of concept overlap. In this work, however, we focus on the model and only observe that matching tends to derive correspondences between attributes of the same concept.

3 Towards Process Matching Networks

Having discussed schema matching networks, we elaborate on how this model may be utilised to explore business processes within a business network. A business process comprises activities that are executed in coordination to achieve a business goal [5]. There is a spectrum of approaches to capture business processes, ranging from informal texts to full-fledged process modelling languages. Against the background of exploring processes within a business network, we capture the essential steps of a process, i.e., the activities. Further details in terms of execution semantics, however, are of interest only once operational commonalities and differences have been judged. Then, behavioural consistency and similarity measures [6,7] allow for conducting a fine-granular analysis.

We propose to rely on a model of a process matching network, which comprises abstract process descriptions given as activity graphs. Such a graph defines a process as a composition of activities, in which abstract activities are composed of more detailed activities. Thus, the model accommodates different level of granularity when capturing a business process. Figure 3 illustrates the model for a financing process. Again, the question arises whether there are reusable concepts shared by network members. Activities in processes often represent an action that is performed on a business object, aka artefact. Hence, there are sets of semantically related activities that relate to the same object and, implicitly, hint at its object lifecycle. This suggests to structure reuse within

a process matching network around these objects. Recent work on part of speech tagging of activity labels [8] helps to identify an action and an object for an activity. Guided by concept libraries like the MIT Process Handbook[3], a process description may be decomposed into concepts, sets of activities that relate to the same business object. For instance, Fig. 3 highlights activities related to the 'Contract' object.

A process description is matched to other descriptions within a business network. This step directly follows the work on schema matching networks, but establishes correspondences between activities instead of schema attributes. Again, concepts guide and structure the matching process.

We foresee different ways of exploring business networks using a process matching network. First, companies with a large functional overlap may be identified. This provides insights on competitors, but also points to potential synergies and suggests collaboration. Second, a company that needs to implement a new business process, e.g., because of changed regulations, may identify companies that implemented the process already. As such, one may benefit from the expertise of companies that have adopted a change. Those may, in turn, sell their knowledge. Third, building a consortium of companies is supported by identifying which companies implement complementary functionality.

4 Conclusions & Outlook

We outlined how the business processes shared within a business network may be explored. Inspired by work on schema matching networks, we proposed to rely on process matching networks, in which abstract process descriptions are matched based on concepts. Those group activities that relate to the same business object.

The outlined approach is a vision and various steps need to be taken to realise it. It needs techniques (1) for extracting and filtering concepts from existing process libraries or reference models, (2) for matching process description in a concept-aware manner, and (3) for querying a process matching network.

References

1. Halevy, A.Y., Ives, Z.G., Madhavan, J., Mork, P., Suciu, D., Tatarinov, I.: The piazza peer data management system. IEEE Trans. Knowl. Data Eng. **16**(7) (2004) 787–798
2. Saha, B., Stanoi, I., Clarkson, K.L.: Schema covering: a step towards enabling reuse in information integration. In ICDE, IEEE (2010) 285–296
3. Bellahsene, Z., Bonifati, A., Rahm, E., eds.: Schema Matching and Mapping. Springer (2011)
4. Smith, K.P., Morse, M., Mork, P., Li, M.H., Rosenthal, A., Allen, D., Seligman, L.: The role of schema matching in large enterprises. In CIDR, www.crdrdb.org (2009)
5. Weske, M.: Business Process Management. Springer (2007)
6. Dijkman, R.M., Dumas, M., van Dongen, B.F., Käärik, R., Mendling, J.: Similarity of business process models: Metrics and evaluation. Inf. Syst. **36**(2) (2011) 498–516
7. Weidlich, M., Mendling, J., Weske, M.: Efficient consistency measurement based on behavioral profiles of process models. IEEE Trans. Software Eng. **37**(3) (2011) 410–429
8. Leopold, H., Smirnov, S., Mendling, J.: On the refactoring of activity labels in business process models. Inf. Syst. **37**(5) (2012) 443–459

[3] http://ccs.mit.edu/ph/