# Exploiting Metrics for Similarity-based Semantic Web Service Discovery

Stefan Dietze, Alessio Gugliotta, John Domingue
*Knowledge Media Institute, The Open University,*
*MK7 6AA, Milton Keynes, UK*
*{s.dietze, a.gugliotta, j.b.domingue}@open.ac.uk*

***Abstract*—Semantic Web Services (SWS) aim at the automated discovery and orchestration of Web services on the basis of comprehensive, machine-interpretable semantic descriptions. However, heterogeneities between distinct SWS representations pose strong limitations w.r.t. interoperability and reusability. Hence, semantic level mediation, i.e. mediation between concurrent semantic representations, is a key requirement to allow SWS matchmaking algorithms to compare capabilities of distinct SWS. In that, semantic level mediation requires to identify similarities across distinct SWS representations. Since current approaches to mediate between distinct service annotations rely either on manual one-to-one mappings or on semi-automatic mappings based on the exploitation of linguistic or structural similarities, these are perceived to be costly and error-prone. We propose a mediation approach enabling the implicit representation of similarities across distinct SWS by grounding these in so-called Mediation Spaces (MS). Given a set of SWS and their respective MS grounding, a general-purpose mediator automatically computes similarities to identify the most appropriate SWS for a given request. A prototypical application illustrates our approach.**

***Keywords-Semantic Web Services; Interoperability; Conceptual Spaces; Discovery; Mediation.***

## I. INTRODUCTION

The increasing availability of a broad variety of Web services raises the need to automatically discover and orchestrate appropriate services for a given need. *Semantic Web Services (SWS)* [9] aim at addressing this challenge on the basis of comprehensive, machine-interpretable semantic descriptions. However, since Web services usually are provided by distinct and independent parties, the actual Web service interfaces as well as their semantic representations are highly heterogeneous. This strongly limits the interoperability and raises the need of mediating between SWS descriptions as well as the actual Web services. Despite the importance of mediation for widespread dissemination of SWS technologies, approaches to mediation are still limited and underdeveloped [19].

In this paper, we particularly address *semantic level mediation* which refers to the resolution of heterogeneities between concurrent semantic representations of services – the actual SWS descriptions – as opposed to data-level mediation, i.e. mediation of the structure, values or formats of input and output (I/O) messages. Therefore, semantic level mediation is particularly important to support the Web service discovery problem.

We argue that semantic level mediation strongly relies on identifying *semantic similarities* between entities across different SWS ontologies [18][26]. However, semantic similarity is not an implicit notion within existing SWS representations (e.g. based on WSMO[1] [9] or OWL-S[2]). Moreover, automatic similarity detection as demanded by semantic mediation requires semantic meaningfulness. But the symbolic approach – i.e. describing symbols by using other symbols without a grounding in the real world – of established SWS representations does not fully entail semantic meaningfulness, since meaning requires both the definition of a terminology in terms of a logical structure (using symbols) and grounding of symbols to a conceptual level [6]. Current approaches to mediation usually foresee the manual development of rather ad-hoc one-to-one mappings or the application of ontology mapping methodologies, mostly based on identifying (a) linguistic commonalities and/or (b) structural similarities [16][4]. Since manually or semi-automatically defining similarity relationships is costly, current approaches are thus not capable to support SWS discovery on a web scale.

In our work, we investigate a similarity-based mediation mechanism in order to overcome the need for manual or semi-automatic formalisations of one-to-one mappings between distinct SWS representations. In this respect, we propose a general purpose mediation approach consisting of (a) a representational approach allowing to implicitly represent similarities and (b) a general-purpose mediator for semantic level mediation, exploiting similarities as represented through (a). In particular, we introduce the concept of *Mediation Spaces (MS)* to enable the implicit representation of semantic similarities across heterogeneous SWS representations through grounding of SWS descriptions into vector spaces. We demonstrate that refining heterogeneous SWS descriptions in multiple shared MS supports similarity-based mediation at the semantic level and implicitly facilitates Web service discovery. The provided general-purpose mediator – implemented as a dedicated mediation Web service – is deployable for any semantic level mediation scenario when

[1] http://www.wsmo.org/2004/d2/v1.0/
[2] http://www.daml.org/services/owl-s/1.0/.

being used together with our proposed representational approach.

The remainder of the paper is organized as follows: Section 2 introduces the SWS mediation problem, while our approach to mediation based on conceptually grounded SWS ontologies is proposed in Section 3. In Section 4, we introduce the implementation of a generic mediator and its deployment in a proof-of-concept application in Section 5. Finally, we discuss and conclude our work in Section 6.

## II. SEMANTIC WEB SERVICES MEDIATION

Before formally introducing the SWS mediation problem, we report below the abstract definitions of SWS and SWS mediation as used throughout the remainder of the paper, together with background information on current mediation approaches.

**Semantic Web Services:** a SWS description (either the description of the Web service or the description of the service request) is formally represented within a particular ontology that complies with a certain SWS reference model. By applying a common formalisation of an ontology [8] to SWS, we define a populated *service ontology* $O$ – as utilised by a particular SWS representation – as a tuple:

$$O = \{C, I, P, R, A\} \subset SWS$$

With $C$ being a set of *n concepts w*here each concept $C_i$ is described through *l(i) concept properties pc*, i.e.:

$$PC_i = \{(pc_{i1}, pc_{i2}, ..., pc_{l(i)}) | pc_{ix} \in C_i\}.$$

$I$ represents all *m instances* where each instance $I_{ij}$ represents a particular instance of a concept $C_j$ and consists of *l(i) instantiated properties pi* instantiating the concept properties of $C_j$:

$$PI_{ij} = \{(pi_{ij1}, pi_{ij2}, ..., pi_{l(i)}) | pi_{ijx} \in I_{ij}\}.$$

Hence, the properties $P$ of an ontology $O$ represent the union of all concept properties $PC$ and instantiated properties $PI$ of $O$:

$$P = \{(PC_1, PC_2, ..., PC_n) \cup (PI_1, PI_2, ..., PI_m)\}$$

Given these definitions, we would like to point out that properties here exclusively refer to so-called data type properties. Hence, we define properties as being distinctive to relations $R$. The latter describe relations between concepts and instances. In addition, $A$ represents a set of *axioms* which define constraints on the other introduced notions. Since certain parts of a SWS ontology describe certain aspects of the Web service (request), such as its capability *Cap*, interface *If* or non-functional properties *Nfp* [5], a SWS ontology can be perceived as a conjunction of ontological subsets:

$$Cap \cup If \cup Nfp = O \subset SWS$$

The semantic capability description, as central element of a SWS description, consists of further subsets, describing the assumptions *As*, effects *Ef*, preconditions *Pre* and postconditions *Post* of a Web Service. However, given the lack of a clear distinction between assumption/effect and pre-/postcondition, we prefer the exclusive usage of assumptions/effects:

$$As \cup Ef = Cap \subset O \subset SWS$$

**SWS mediation:** mediation aims at addressing heterogeneities among distinct SWS to support all stages that occur at SWS runtime, namely *discovery*, *orchestration* and *invocation*. In contrast to [19][5], we classify the mediation problem into (i) *semantic level* and (ii) *data level mediation* (Figure 1). The following simplified picture illustrates the chronological order of different mediation tasks at SWS runtime.

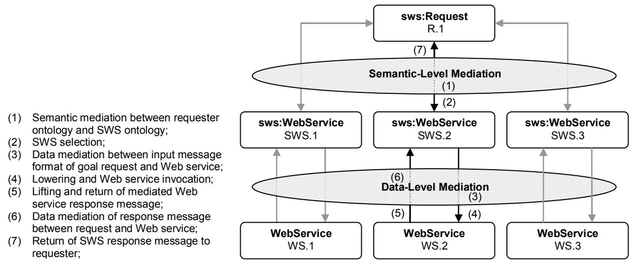


Figure 1. Semantic level and data level mediation as part of SWS discovery, orchestration and invocation.

Whereas (i) refers to the resolution of heterogeneities between concurrent semantic representations of services – e.g. by aligning distinct SWS representations – (ii) refers to the mediation between mismatches related to the Web service implementations themselves, i.e. related to the structure, value or format of I/O messages. Hence, semantic level mediation primarily supports the discovery stage, whereas data level mediation occurs during orchestration and invocation. Please note that, for the sake of simplification, Figure 1 just depicts mediation between a SWS request and multiple SWS, while leaving aside mediation between different SWS or between different requests.

Several approaches, such as [1][2][15][20][23][26], aim at addressing the mediation issue partially by dealing either with (i) or (ii). For instance, [2] proposes a semantic mediation framework for scientific workflows relying on the notion of semantic type and structural type, defined in a shared ontology. The semantic type gives a meaning to data, and the structural type is the data schema. As in [23] their work adapts data with a common semantic type but different structural types. In contrast, [26] provides an attempt to support similarity detection for mediation within SWS composition by exploiting syntactic similarities between SWS representations. However, it can be stated that all the above mentioned approaches rely on the definition of a priori mappings, the agreement of a shared ontology or the exploitation of semi-automatic ontology mapping approaches. Hence, providing a generic solution to mediation between heterogeneous SWS remains a central challenge.

**Semantic level mediation as a similarity computation problem:** In this paper, we exclusively address semantic level mediation, what is perceived to be a fundamental requirement to further exploit SWS approaches on a Web scale. To understand the needs of semantic level mediation, it is necessary to understand the requirements of the SWS discovery task to which semantic level mediation is supposed to contribute. In order to identify whether a particular SWS $S_1$ is potentially relevant for a given request $S_2$, a SWS broker has to compare the capabilities of $S_1$ and $S_2$, i.e. it has to identify whether the following holds true:

$$As_2 \subset As_1 \cup Ef_2 \subset Ef_1$$

However, in order to compare distinct capabilities of available SWS which each utilize a distinct vocabulary, these vocabularies have to be aligned. For instance, to compare whether an assumption expression $As_1 \equiv \neg I_1 \cup I_2$ of one particular $SWS_1$ is the same as $As_2 \equiv I_3 \cup \neg I_4$ of another $SWS_2$, where $I_i$ represents a particular instance, matchmaking engines have to perform two steps: (a) identification of relationships between concepts/instances involved in distinct SWS representations; (b) evaluation whether the semantics of the two SWS expressions match each other. Whereas current SWS execution environments exclusively focus on (b), semantic level mediation also requires mediation between different ontologies, as in (a), and can be perceived as a particular instantiation of the *ontology mapping* problem [26]. The goal is, to establish formal relations between a set of knowledge entities $E_1$ from an ontology $O_1$ – used to represent a particular SWS $S_1$ – with entities $E_2$ which represent the same or a similar semantic meaning in a distinct ontology $O_2$ [4][8] which is used to represent an additional SWS $S_2$. The term *set of entities* here refers to the union of all concepts $C$, instances $I$, relations $R$ and axioms $A$ defined in a particular SWS ontology. In that, semantic mediation strongly relies on identifying *semantic similarities* [1] between entities across different SWS ontologies. Hence, the identification of similarities is a necessary requirement to solve the mediation problem for multiple heterogeneous SWS representations [18][26][7]. However, in this respect, the following issues have to be taken into account:

*I1 - Symbolic SWS representations lack meaningfulness and are ambiguous*: similarity-detection across distinct SWS representations requires semantic meaningfulness which inherently represents semantic similarity between represented entities. However, the symbolic approach, i.e. describing symbols by using other symbols, without a grounding in the real world, of established SWS representation standards, leads to ambiguity issues and does not fully entail semantic meaningfulness, since meaning requires both the definition of a terminology in terms of a logical structure (using symbols) and grounding of symbols to a conceptual level [6][15].

*I2 - Lack of automated similarity-detection methodologies*: Describing the complex notion of specific SWS capabilities in all their facets is a costly task and may never reach semantic completeness due to *I1*. While capability representations across distinct SWS representations – even those representing the same real-world entities – hardly equal another, semantic similarity is not an implicit notion within SWS representations. But manually or semi-automatically defining similarity relationships is costly. Moreover, such relationships are hard to maintain in the longer term.

Given the lack of inherent similarity representation, current approaches to ontology mapping could be applied to facilitate SWS mediation. These approaches aim at semi-automatic similarity detection across ontologies mostly based on identifying linguistic and/or structural similarities between entities of distinct ontologies [16][4][12]. Work following a combination of such approaches in the field of ontology mapping is reported in [13][11][16][6]. However, it can be stated, that such approaches require manual intervention, are costly and error-prone, and hence, similarity-computation remains as central challenge.

In our vision, instead of semi-automatically formalising individual mappings, methodologies to automatically compute or implicitly represent similarities across distinct SWS representations are better suited to facilitate SWS mediation.

## III. SIMILARITY-BASED SWS DISCOVERY BASED ON MEDIATION SPACES

To overcome the issues introduced in the previous section, we propose a mediation approach which utilises a novel representation mechanism which implicitly represents similarities.

### *A. Mediation Spaces for SWS*

We propose a representational approach which grounds a SWS representation in so-called *Mediation Spaces (MS),* which are inspired by *Conceptual Spaces (CS)* [10] and enable the implicit representation of semantic similarities across heterogeneous SWS representations provided by distinct agents. MS propose the representation of concepts which are used as part of SWS descriptions as multidimensional geometrical vector spaces which are defined through sets of quality dimensions. Instances are represented as vectors (members), i.e. particular points in a MS where similarity between two vectors is indicated by their spatial distance. Hence, refining heterogeneous SWS descriptions into multiple shared MS supports similarity based mediation at the semantic level and consequently facilitates SWS discovery.

Whereas CS allow the representation of semantic similarity as a notion implicit to a constructed knowledge model, it can be argued, that representing an entire SWS through a coherent MS might not be feasible, particularly when attempting to maintain the meaningfulness of the spatial distance as a similarity measure.

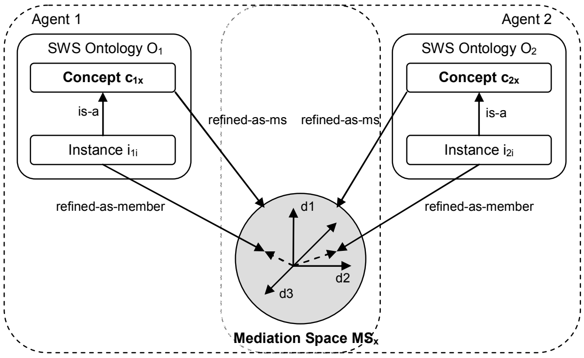


Figure 2. Representing heterogeneous SWS representations through shared Mediation Spaces.

Therefore, we claim that MS are a particularly promising model when being applied to individual concepts – as part of SWS descriptions – instead of representing an entire SWS ontology in a single MS. In that, we would like to highlight that we consider the representation of a set of *n* concepts *C* of a SWS ontology *O* through a set of *n* MS (Figure 2). Hence, instances of concepts are represented as members (i.e. vectors) in the respective MS. While still taking advantage from implicit similarity information within a MS, our hybrid approach – combining SWS descriptions with multiple MS – allows to overcome CS-related issues, such as the lack of expressivity for arbitrary relations, by maintaining the advantages of ontology-based SWS representations. Please note that our approach relies on the agreement on a common set of MS for a given set of distinct SWS ontologies, instead of a common agreement on the used ontologies/vocabularies themselves. Thus, whereas in the latter case two agents have to agree on a common ontology at the concept and instance level, our approach requires just agreement at the concept level, since instance similarity becomes an implicit notion. Moreover, we assume that the agreement on ontologies at the concept level (Figure 2) becomes an increasingly widespread case, due to, on the one hand, increasing use of upper-level ontologies such as DOLCE[3], SUMO[4] or OpenCyc[5] which support a certain degree of commonality between distinct ontologies. On the other hand, SWS ontologies often are provided within closed environments, for instance, virtual organisations, where a common agreement to a certain extent is ensured. In such cases, the derivation of a set of common MS is particularly applicable and straightforward.

In order to refine and represent SWS descriptions within a MS, we formalised the MS model into an ontology, currently being represented through OCML [14]. The ontology enables the instantiation of a set of MS to represent a given set of concepts as part of SWS descriptions. Referring to [19], we formalise a MS as a vector space defined through quality dimensions $d_i$ of *MS*. Each dimension is associated with a certain metric scale, e.g. ratio, interval or ordinal scale. To reflect the impact of a specific quality dimension on the entire MS, we consider a prominence value *p* for each dimension [19]. Therefore, a MS is defined by

$$MS^n = \{(p_1 d_1, p_2 d_2, ..., p_n d_n) | d_i \in MS, p_i \in \Re\}.$$

However, usage context, purpose and domain of a particular MS strongly influence the ranking of its quality dimensions. This clearly supports our position of describing distinct MS explicitly for individual concepts. Please note that we enable dimensions to be detailed further in terms of subspaces. Hence, a dimension within one MS may be defined through another MS by using further dimensions. In such a case, the particular quality dimension $d_j$ is described by a set of further quality dimensions. In this way, a MS may be composed of several subspaces and consequently, the description granularity can be refined gradually. Furthermore, dimensions may be correlated. Information about correlation is expressed through axioms related to a specific quality dimension instance.

A member *M* – representing a particular instance – of the MS is described through a set of valued dimension vectors $v_i$:

$$M^n = \{(v_1, v_2, ..., v_n) | v_i \in M\}$$

With respect to [19], we define the semantic similarity between two members of a space as a function of the Euclidean distance between the points representing each of the members. However, we would like to point out that different distance metrics could be considered, dependent on the nature and purpose of the MS. Given a MS definition *MS* and two members *V* and *U*, defined by vectors $v_0$, $v_1$, ..., $v_n$ and $u_1$, $u_2$, ..., $u_n$ within *MS*, the distance between *V* and *U* can be calculated as:

$$dist(u,v) = \sqrt{\sum_{i=1}^{n} p_i \left( \left( \frac{u_i - \bar{u}}{s_u} \right) - \left( \frac{v_i - \bar{v}}{s_v} \right) \right)^2}$$

where $\bar{u}$ is the mean of a dataset *U* and $s_u$ is the standard deviation from *U*. The formula above already considers the so-called Z-transformation or standardization [24] which facilitates the standardization of distinct measurement scales utilised by different quality dimensions in order to enable the calculation of distances in a multi-dimensional and multi-metric space.

3 http://www.loa-cnr.it/DOLCE.html
4 http://www.ontologyportal.org/
5 http://www.opencyc.org/

### B. Aligning SWS Capabilities through Mediation Spaces

Following our vision, the provisioning of SWS representations is a highly heterogeneous and distributed procedure that is accomplished autonomously by distinct agents. In particular, we distinguish two groups of involved agents: (C1) *distributed SWS providers and consumers* and (C2) *centralised SWS maintainers*. The existence of C2 is implied by the broker-based nature of SWS technologies.

Specifically, the overall procedure of providing SWS following our approach is based on the following steps:

- S1. Provisioning of a central SWS runtime environment (C2).
- S2. Provisioning of SWS representations $S^n$ (C1).
- S3. Providing appropriate $MS_i$ for each distinct real-world entity represented within an available SWS ontology $O$.
  - S3.1. Representing concept properties $pc_{ij}$ of $C_i$ as dimensions $d_{ij}$ of $MS_i$ (C2).
  - S3.2. Assignment of metrics to each quality dimension $d_{ij}$ (C2).
  - S3.3. Assignment of prominence values $p_{ij}$ to each quality dimension $d_{ij}$ (C2).
  - S3.4. Representing all instances $I_{ik}$ of $C_i$ as members in $MS_i$ (C1).

Whereas *S1* and *S2* are foreseen within the SWS vision in general, *S3* represents an additional activity aiming at providing the representational facilities required to realise our mediation approach. Please note that certain steps are performed by a centralised SWS maintainer (C2) – such as the provisioning of the SWS environment (*S1*) and the representation of concepts involved in SWS descriptions as MS (*S3.1 – S3.3*) – whereas others are accomplished by distributed Web service providers (C1) – such as the provisioning of SWS descriptions (*S2*) and the representation of instances as members following the defined MS (*S3.4*). In that, this methodology takes into account the fact that Web services as well as their semantic annotations usually are provided by distributed and independent actors.

## IV. A SIMILARITY-BASED MEDIATION SERVICE

To facilitate our MS-based approach, we provided a general-purpose mediator – implemented as a particular mediation service – which in fact is composed of two standard Web services (MWS.1, MWS.2). Given the ontological refinement of SWS descriptions into MS as introduced in Section III, the mediation service is reusable and can be deployed to solve all sorts of semantic level mediation scenarios.

At runtime, the first MWS.1 is invoked. Its inputs are a particular $SWS_i$ (e.g. a service request description), named *base,* and the SWS descriptions of all $x$ available services that are potentially relevant for the base*:*

$$in(MWS.1) = SWS_i \cup \{SWS_1, SWS_2, ..., SWS_x\}$$

Each SWS contains a set of concepts $C=\{c_1..c_m\}$ and instances $I=\{i_1..i_n\}$. Exchange of such ontological descriptions through SOAP is enabled by using an XML-serialisation as exchange format. MWS.1 first identifies all members $M(SWS_i)$ – in the form of valued vectors $\{v_1..v_n\}$ – refining the instance $i_l$ of the base as proposed in Section III. In addition, for each concept $c$ within the base the corresponding mediation space representations $MS=\{MS_1..MS_m\}$ are retrieved. Similarly, for each $SWS_j$ related to the base, members $M(SWS_j)$ – which refine capabilities of $SWS_j$ and are represented in one of the mediation spaces $MS_1..MS_m$, – are retrieved. In that, the output of MWS.1 represents also the input of MWS.2 and can be described as follows:

$$in(MWS.2) = MS \cup M(SWS_i) \cup \{M(SWS_1), M(SWS_2), ..., M(SWS_x)\}$$

MWS.2 aims at computing the semantic similarities between the capability descriptions of $SWS_i$ and the $x$ associated $SWS_j$. In order to do so, MWS.2 is provided with the retrieved ontological descriptions, namely all members $M(SWS_i)$ and $M(SWS_j)$ and the respective space definitions $MS$. Based on the ontological descriptions of the input, for each member $v_l$ within $M(SWS_i)$, MWS.2 computes the Euclidean distances to any member of all $M(SWS_j)$ which is represented in the same space $MS_j$ as $v_l$. In case one set of members $M(SWS_j)$ contains several members in the same MS – e.g. $SWS_j$ targets several instances of the same kind – the algorithm just considers the closest distance since the closest match determines the appropriateness for a given goal. For example, if one SWS supports several different locations, just the one which is closest to the one required by $SWS_i$ determines the appropriateness.

Consequently, mediation service MWS.2 computes a set of $x$ sets of distances $Dist(SWS_i)=\{Dist(SWS_i,SWS_1), Dist(SWS_i,SWS_2) .. Dist(SWS_i,SWS_x)\}$ where each $Dist(SWS_i,SWS_j)$ contains a set of distances $\{dist_1..dist_n\}$ where any $dist_i$ represents the distance between one particular member $v_i$ of $SWS_i$ and one member refining one instance of the capabilities of $SWS_j$. Hence, the overall similarity between the base $SWS_i$ and any $SWS_j$ could be defined as being reciprocal to the mean value of the individual distances between all instances of their respective capability descriptions and hence, is calculated as follows:

$$Sim(SWS_i, SWS_j) = \left(\overline{Dist(SWS_i, SWS_j)}\right)^{-1} = \left( \frac{\sum_{k=1}^{n} (dist_k)}{n} \right)^{-1}$$

The final output of the composed mediator is a set of $x$ similarity values – computed as described above – which each indicates the similarity between the base $SWS_i$ and one of the $x$ target SWS:

$$out(SWS.1.2) = \{Sim(SWS_i, SWS_1), Sim(SWS_i, SWS_2), .., Sim(SWS_i, SWS_x)\}$$

As a result, the most similar $SWS_j$, i.e. the closest associated SWS, can be invoked. In order to ensure a certain degree of overlap between the actual request and the invoked functionality, we also defined a *threshold similarity value T* which determines the similarity threshold for any potential invocation.

## V. DEPLOYING SIMILARITY-BASED MEDIATION BETWEEN WEATHER FORECAST SERVICES

Even though our approach could be applied to any kind of SWS reference model, we adopted WSMO [9] to implement a proof-of-concept prototype. Particularly, we deployed the mediation Web services introduced in the previous section as WSMO mediator.

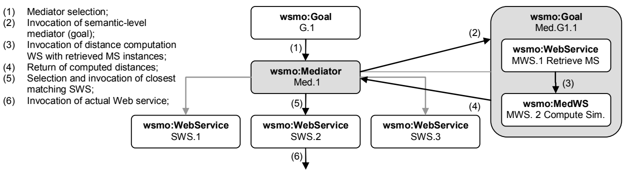


Figure 3. Semantic level mediation facilitated through a general-purpose WSMO mediator.

Moreover, we make use of IRS-III [3], a WSMO-compliant reasoner and SWS broker environment. As example, Figure 3 illustrates the functionality of our mediator being deployed to mediate between a goal request and several WSMO SWS. In this example scenario, the WSMO mediator (*Med.1*) mediates between a given goal *G.1* and a set of 3 potentially relevant Web services (*SWS.1, SWS.2, SWS.3*). According to WSMO specifications, *Med.1* is associated with a distinct goal (*G.1.1*) that, in our case, is achieved by the orchestration of MWS.1 and MWS.2 (Section IV). In this example, similarity-based mediation is applied during SWS discovery.

The general schema depicted in Figure 3 has been also actualised within an initial proof-of-concept prototype application which mediates between different weather forecast Web services. Here, $SWS_1$, $SWS_2$ and $SWS_3$ provide weather forecast information for different locations. Each service has distinct constraints, and thus distinct SWS descriptions. In detail, $SWS_1$ is able to provide forecasts for France and Spain while $SWS_2$ and $SWS_3$ are providing forecasts for the United Kingdom. All services show different Quality of Service (QoS) parameters. Three distinct service ontologies – $O_1$, $O_2$, and $O_3$ – together with a SWS request ontology $O_4$ had been created, each defining the capability of the respective service by using distinct vocabularies. For example, $SWS_2$ considers concepts representing the notions of location and QoS together with corresponding instances (see also Table 1):

$$\{(country, QoS), (UK, QoS2)\} \subset O_2 \subset SWS_2$$

By applying the representational approach proposed in Section III, each concept of the involved heterogeneous SWS representations had been refined as a shared MS, while instances - defining the capabilities of available SWS and SWS requests - were defined as members. No explicit relations were formalised across ontologies. Instead, similarities between instances are computed by means of distance calculation in the shared MS.

For example, a simplified space (*$MS_1$: Location Space* in Figure 4) was utilized to refine geographical notions (e.g. country) by using two dimensions indicating the geospatial position of the location:

$$\{(p_1 l_1, p_2 l_2)\} = \{(latitude, longitude)\} = MS_1$$

The two dimensions latitude and longitude are equally ranked, and hence, a prominence value of 1 has been applied to each dimension. Note that each of the depicted concepts and instances, such as *$O_2$:UK* and *$O_3$:UK*, are distinct and independent from each other, and thus might show heterogeneities, such as distinct labels, for instance *United Kingdom* and *Great Britain*. In the case of *$O_2$:UK* and *$O_3$:UK,* these two instances are refined by two distinct members:

$L_1(SWS_2) = \{(v_1 = 55.378051, v_2 = \text{-}3.435973) | v_i \in MS_1\}$ and

$L_1(SWS_3) = \{(v_1 = 55.378048, v_2 = \text{-}3.435963) | v_i \in MS_1\}$. Each member has been defined by different individuals applying similar, but non-equivalent geodata.

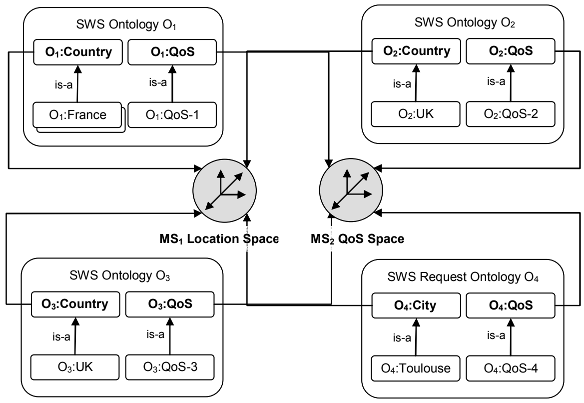


Figure 4. Grounding assumptions of distinct weather forecast SWS to common MS.

In addition, a second space (*$MS_2$: QoS Space* in Figure 4) has been defined by three dimensions – latency (in ms), throughput (number of Web services), availability (in %):

$$\{(p_1 r_1, p_2 r_2, p_3 r_3)\} = \{(latency, throughput, availability)\} = MS_2$$

In that, assumptions of available SWS had been described independently in terms of simple conjunctions of instances which were individually refined in shared MS as shown in Table 1. Potential service consumers define a request as a WSMO goal (e.g. $SWS_4$ in Figure 4) together with the set of input parameters and the underlying assumptions. Analogous to the SWS descriptions, instances used to define the goal assumptions are grounded to members in the corresponding MS.

TABLE 1. ASSUMPTIONS OF INVOLVED SWS AND SWS REQUESTS DESCRIBED IN TERMS OF VECTORS IN $MS_1$ AND $MS_2$.

| | **Assumption** $Ass_{SWSi} = (L_{1SWSi} \cup L_{2SWSi} \cup .. \cup L_{nSWSi}) \cup (Q_{1SWSi} \cup Q_{2SWSi} \cup .. \cup Q_{mSWSi})$ | |
|---|---|---|
| | Members $L_i$ in $MS_1$ (locations) | Members $C_j$ in $MS_2$ (QoS) |
| $SWS_1$ | $L_{1(SWS1)}$={(46.227644, 2.213755)}<br>$L_{2(SWS1)}$={(40.463667, -3.74922)} | $Q_{1(SWS1)}$={(155, 2, 91)} |
| $SWS_2$ | $L_{1(SWS2)}$={(55.378051, -3.435973)} | $Q_{1(SWS2)}$={(15, 50, 98)} |
| $SWS_3$ | $L_{1(SWS3)}$={(55.378048, -3.435963)} | $Q_{1(SWS3)}$={(78, 5, 95)} |
| $SWS_4$ | $L_{1(SWS4}$={(55.378048, -3.435963)} | $Q_{1(SWS4)}$={(0,100,100)} |

As shown in Table 1, the request $SWS_4$ assumes a SWS which provides weather forecast for the location UK ($L_1(SWS_4)$) and ideal QoS ($Q_1(SWS_4)$) demanding zero latency but high throughput and availability. Though no exact SWS matches these criteria, at runtime similarities are calculated between $SWS_4$ and the related SWS ($SWS_1$, $SWS_2$, $SWS_3$) through the mediation services introduced in Section IV. This led to the calculation of the following similarity values:

TABLE 2. AUTOMATICALLY COMPUTED SIMILARITIES BETWEEN SWS REQUEST $SWS_4$ AND AVAILABLE SWS.

| | Similarities |
|---|---|
| $SWS_1$ | 0.010290349 |
| $SWS_2$ | 0.038284954 |
| $SWS_3$ | 0.016257476 |

Given these similarities, our mediation service automatically selects the most similar SWS ($SWS_2$) and triggers its invocation, potentially leading to further data level mediation tasks.

## VI. DISCUSSION AND CONCLUSIONS

In order to facilitate SWS interoperability we proposed a semantic mediation approach based on two contributions: (a) a hybrid representation using a combination of symbolic SWS representations and concept groundings in multiple MS and (b) a general-purpose mediation service enabling to compute similarities between distinct SWS representations. MS follow the vector space theory and enable the representation of instances as vectors to facilitate the automatic computation of similarities by means of spatial distances. A dedicated MS formalisation enables the instantiation of a corresponding MS (member) for each individual concept (instance) of any arbitrary SWS ontology.

The introduced two-fold representational approach supports implicit representation of similarities between instances across heterogeneous SWS, and consequently, provides a means to facilitate Web service interoperability. In fact, given the set of SWS representations grounded into MS, our general-purpose mediation Web service is able to compute their similarities in order to identify the best possible match. Furthermore, our approach is supported by a formal method on how to derive MS representations for individual concepts of arbitrary SWS representations. To evaluate our approach, we deployed a prototypical application based on WSMO in a weather forecast booking scenario.

The proposed approach has the potential to significantly reduce the effort required to mediate between distinct heterogeneous SWS and the extent to which distinct parties have to share their conceptualisations. Whereas traditional mediation methodologies rely on either manual formalisations of one-to-one mappings or mechanisms to semi-automatically detect similarities at the concept and the instance level, our approach supports automatic similarity-computation between instances though requiring a common agreement on a shared MS. However, even for the case of heterogeneous MS, traditional semi-automatic mapping methodologies could be applied to initially align distinct MS. In addition, incomplete similarities are computable between partially overlapping MS. Given the nature of our approach – aiming at mediating between sets of concepts/instances which are used to annotate particular SWS – we argue that our solution is particularly applicable to SWS frameworks which are based on rather light-weight service semantics such as WSMO-Lite [25], SAWSDL[6] or OWL-S. Moreover, by representing SWS through vectors which are independent from the underlying representation language, potentially our approach could bridge between concurrent SWS reference models.

However, the authors are aware that a considerable amount of additional effort is required to establish MS-based representations. Future work will investigate the scalability of our approach, as well as its reusability in distinct application domains. Moreover, while overcoming issues introduced in Section II, further issues remain. For example, whereas defining instances, i.e. vectors, within a given MS appears to be a straightforward process of assigning specific quantitative values to quality dimensions, the definition of the MS itself is not trivial and dependent on individual perspectives and subjective appraisals. In addition, whereas semantics of instances are grounded to metrics within a MS, the quality dimensions themselves are subject to ones interpretation what might lead to ambiguity issues. Nevertheless, distance calculation relies on the fact that resources are described in equivalent geometrical spaces. However, particularly with respect to the latter, traditional ontology and schema matching methods could be applied to align heterogeneous spaces. In addition, we would like to point out that the increasing usage of upper level ontologies such as DOLCE or SUMO and the progressive reuse of ontologies, particularly in loosely coupled organisational environments, leads to an increased sharing of (SWS) formalisations, particularly at the schema level. As a result, our proposed hybrid

[6] http://www.w3.org/2002/ws/sawsdl/

representational model and mediation approach becomes increasingly applicable by further enabling similarity-computation at the instance-level towards the vision of interoperable Web services.

# REFERENCES


[1] Bicer, V., Kilic, O., Dogac, A., Laleci, G.B. Archetype-based semantic interoperability of web service messages in the health care domain. International Journal of Semantic Web and Information Systems, 1(4):1–23, October 2005.

[2] Bowers, S., Ludäscher, B. An ontology-driven framework for data transformation in scientific workflows. In E. Rahm, editor, DILS, volume 2994 of Lecture Notes in Computer Science, pages 1–16. Springer, 2004.

[3] Cabral, L., Domingue, J., Galizia, S., Gugliotta, A., Norton, B., Tanasescu, V., Pedrinaci, C.: IRS-III: A Broker for Semantic Web Services based Applications. In proceedings of the 5th International Semantic Web Conference, Athens, USA (2006).

[4] Choi, N., Song, I., and Han, H. (2006), A survey on ontology mapping, SIGMOD Rec., Vol. 35, No. 3. (September 2006), pp. 34-41.

[5] Cimpian, E., Mocan, A., Stollberg, M. (2006) Mediation Enabled Semantic Web Services Usage. Proceedings of the 1st Asian Semantic Web Conference (ASWC 2006), September 6, 2006, Beijing, China, 2006.

[6] Cregan, A. (2007), Symbol Grounding for the Semantic Web. 4th European Semantic Web Conference 2007, Innsbruck, Austria.

[7] Dietze, S., Gugliotta, A., Domingue, J., (2008) Conceptual Situation Spaces for Situation-Driven Processes. 5th European Semantic Web Conference (ESWC), Tenerife, Spain.

[8] Ehrig, M, Sure, Y. (2004), Ontology Mapping - An Integrated Approach, in Proceedings of ESWS, 2004.

[9] Fensel, D., Lausen, H., Polleres, A., de Bruijn, J., Stollberg, M., Roman, D., Domingue, J. (2006): Enabling Semantic Web Services – The Web service Modelling Ontology, Springer 2006.

[10] Gärdenfors, P. (2000), Conceptual Spaces - The Geometry of Thought. MIT Press, 2000.

[11] Giunchiglia, F., Shvaiko, P., and Yatskevich, M. S-Match: An Algorithm and an Implementation of Semantic Matching. ESWS 2004, 61-75

[12] Maedche, A., Staab, S., Measuring Similarity between Ontologies, Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, LNCS Vol. 2473, Springer, pp. 15-21, 2002.

[13] Mitra, P., Noy, F. N., Jaiswals, A. (2005), OMEN: A Probabilistic Ontology Mapping Tool, International Semantic Web Conference 2005.

[14] Motta, E. (1998). An Overview of the OCML Modelling Language, the 8th Workshop on Methods and Languages, 1998.

[15] Mrissa, M., Ghedira, C., Benslimane, D., Maamar, Z., Rosenberg, F., Dustdar, S. A context-based mediation approach to compose semantic web services. ACM Trans. Internet Techn., 8(1), 2007.

[16] Noy, N.F., and Musen, M.A. The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping. International Journal of Human-Computer Studies 59 (2003), 983-1024.

[17] Object Management Group: Business Process Modelling Notation Specification (http://www.omg.org/docs/dtc/06-02-01.pdf).

[18] Qu, Y., Hu, W., Cheng, G. (2006) Constructing Virtual Documents for Ontology Matching, *WWW 2006*, May 23–26, 2006, Edinburgh, Scotland. ACM 1595933239/06/0005.

[19] Paolucci, M., Srinivasan, N., Sycara, K. Expressing WSMO Mediators in OWL-S, in proceedings of Semantic Web Services: in 3rd International Semantic Web Conference (ISWC 2004).

[20] Radetzki, U., Cremers, A. B. Iris: A framework for mediator-based composition of service-oriented software. In ICWS, pages 752–755. IEEE Computer Society, 2004.

[21] Raubal, M. (2004). Formalizing Conceptual Spaces. in: A. Varzi and L. Vieu (Eds.), Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004). Frontiers in Artificial Intelligence and Applications 114, pp. 153-164, IOS Press, Amsterdam, NL.

[22] Schwering, A. (2005). Hybrid Model for Semantic Similarity Measurement, in R. Meersman and Z. Tari (Eds.): CoopIS/DOA/ODBASE 2005, LNCS 3761, pp. 1449 – 1465, 2005..

[23] Spencer, B., Liu, S. Inferring data transformation rules to integrate semantic web services. In S. A. McIlraith, D. Plexousakis, and F. van Harmelen, editors, Int’l Semantic Web Conference, volume 3298 of Lecture Notes in Computer Science, pages 456–470. Springer, 2004.

[24] Suppes, P., D. M. Krantz, et al. (1989). Foundations of Measurement - Geometrical, Threshold, and Probabilistic Representations. San Diego, California, USA, Academic Press, Inc.

[25] Vitvar, T., Kopecký, J., Viskova, J., Fensel, D. WSMO-Lite Annotations for Web Services. ESWC 2008: 674-689.

[26] Wu, Z., Ranabahu, A., Gomadam, K., Sheth, A.P., Miller, J.A., Automatic Composition of Semantic Web Services using Process Mediation, Proceedings of the 9th International Conference on Enterprise Information Systems (ICEIS'07), Funchal, Portugal Jun. 2007. pages 453-461.