Journal of Universal Computer Science, vol. 19, no. 9 (2013), 1315-1333
submitted: 12/2/13, accepted: 26/3/13, appeared: 1/5/13 © *J.UCS*

An Enhanced Process of Concept Alignment for Dealing with Overweight and Obesity

María de Lourdes Martínez-Villaseñor
(Universidad Panamericana Campus México, Mexico City, México
lmartine@up.edu.mx)

Miguel González-Mendoza
(Tecnológico de Monterrey, Campus Estado de México, Atizapán de Zaragoza, México
mgonza@itesm.mx)

Abstract: A major challenge for creating personalized diet and activity applications is to capture static, semi-static and dynamic information about a person in a user-friendly way. Sharing and reusing information between heterogeneous sources like social networking applications, personal health records, specialized applications for diet and exercise monitoring, and personal devices with attached sensors can achieve a better understanding of the user. Gathering distributed user information from heterogeneous sources and making sense of it to enable user model interoperability entails handling the semantic heterogeneity of the user models. In this paper, we enhance the process of concept alignment to automatically determine semantic mapping relations to enable interoperability between heterogeneous health and fitting applications. We add an internal structure similarity measure to increase the quality of generated mappings of our previous work. We show that the addition of an internal structure analysis of source data in the process of concept alignment improves the efficiency and effectiveness of measuring results. Constrain and data type verification done in the internal structure analysis proved to be useful when dealing with common conflicts between concepts.

Keywords: User modeling interoperability, schema matching, overweight and obesity, diet and exercise monitoring
Categories: M.4, M.5, M.8

1 Introduction

According to the World Health Organization (WHO), overweight and obesity are serious public health challenges in the WHO European Region, and their prevalence is growing worldwide even in countries with low or middle income[WHO Regional Office, 07].If a person consumes more calories than she/he expends, she/he will gain weight, resulting in overweight and obesity. This is due to unbalanced diets and low physical activity. Many personal factors must be taken into account when designing personal diet and exercise goals and plans, for example: Basic demographic and anthropometric data, nutrition preferences, allergies and restrictions, medical conditions, activity patterns and preferences, cultural and religious factors.

In brief, static, semi-static and dynamic information is needed to succeed when develop personal goals and plans to deal with overweight and obesity. There are many fitness applications and devices in the market nowadays that offer to help monitoring personal diet and exercise, but each of them requires a lot of effort from the user explicitly capturing basic profile information, preferences, interests and daily records. Transferring data from personal devices that record activity with sensors is also needed. One of the major challenges for creating personalized health applications is to capture the information needed for customizing user-friendly way to avoid having to repeat this task for each application. Ubiquitous computing and ambient intelligence must deliver valuable services "driven with humanistic concerns" [Rui, 10].

Sharing and reusing profile information from health and fitness can prevent the user from repeatedly capturing the same information in several applications and services; helps deal with "cold start" problem of new applications and services; and provides enrichment to existing user models. Gathering distributed user information from heterogeneous sources and making sense of it to enable user model interoperability entails handling the semantic heterogeneity of the user models[Carmagnola, 09]. This heterogeneity is especially serious in the health and fitness domain, given the wide variety of fitness apps, gadgets and activity tracking devices. There is high dynamicity in an open environment like the Web and a domain with total autonomy of the stakeholders. It is not feasible for applications to adhere to the same user model representation. Ubiquitous user model interoperability can be achieved with a mediation-based approach using schema matching techniques.

If we examine data exacted from different applications and devices used to deal with obesity and overweight, we can find conflicts at different levels between similar concepts even hard to decide for a human expert.

Possible conflicts as the ones described by Sosnovsky et al [Sosnovsky, 09] occur when trying to align the source document concepts with the ubiquitous user model. Naming conflicts are frequent when the same label has different meaning in two models or different labels are used for the same concept. Applications also use complex types and different granularity to express the same concept as others, grouping or decomposing data. Labels in these cases are not easily detected as meaningful, for example a date decomposed in single integers with confusing descriptions (*y,m,d,m,s,f*). Lexical, structural and semantic similarity measures that do not consider internal structure constrains are not sufficient to deal with these conflicts. Source documents usually include valuable information like data types and enumerated values that can be used to resolve the mentioned conflicts.

In this paper, we enhance the process of concept alignment to automatically determine semantic mapping relations to enable interoperability between heterogeneous health and fitting applications. We add an internal structure similarity measure to increase the quality of generated mappings of our previous work [Martinez-Villaseñor, 12; Martinez-Villaseñor, 12a]. The addition of this similarity measure in the element level matching phase of the process of concept alignment improves the efficiency and effectiveness of measuring results, helping to prevent mismatches occurred due to the conflicts described above. Final data type verification increases the possibility of interchanging concepts with *exactMatch* semantic relation mapping.

The improvement in the quality of generated mappings supports the integration of profile information of new consumers and suppliers to the ubiquitous user model that enables heterogeneous sources interoperability.

We provide an example to illustrate how reusing and sharing profile information from heterogeneous sources can be done and improved with the mediation of our ubiquitous user model and the process of concept alignment. In this example, we show how information of different profile suppliers can be used to enrich fitness applications to deal with obesity and overweight.

The rest of the paper is organized as follows: in section 2 we present a short survey of the field of ubiquitous user modeling interoperability. We briefly overview of schema matching and schema integration approaches in section 3. We explain the U2MIO ontology and the enhanced process of concept alignment for ubiquitous user model interoperability in section 4.We present experiments and results of sharing and reusing profile information in the scenario of dealing with overweight and obesity in section 5. We conclude an outline our future work in section 6.

2 Ubiquitous User Modeling Interoperability

From literature [Berkovsky, 09; Carmagnola, 11; Viviani, 10], we see that current research in ubiquitous user modeling has two major approaches: (i) standardization based user modeling based in semantic standardization of user model defining a common ontology and language; (ii) mediation-based user modeling using mediation techniques to build semantic bridges between representations.

In standardization-based user modeling approach, sharing and reusing user models is easier, but it requires all systems to adhere to a standardized ontology and representation [Heckmann 05; Kay, 03; Razmerita, 03; Dolog 05; Mehta 05]. Recently, it has become clear that developing a commonly accepted ontology to deal with sparseness of data and heterogeneity of sources in a multi-application environment is not a feasible solution.

As we mentioned in the introduction, this statement is true for highly dynamic environments and domains where applications and devices always change. Nowadays we have ". a whole set of devices placed around us providing users with an intelligent background" [Bravo, 06]. In the health and fitness domain, total autonomy of stakeholders cause high dynamicity as described in [Shvaiko 06]. Fitness application providers are free to decide what data to store and its structure, and are free to change these decisions any time. They design their own transfer mechanisms, policies and devices and change them any time too. New applications and devices appear and become obsolete or outdated, so frequently new profile providers and consumers are willing to participate in the interoperability process. A standardization-based user model approach is not feasible.

The mediation approach consists in mapping different user model representation. Berkovsky et al.[Berkovsky, 09] explained that "Mediation deals with transferring user modeling data from one representation (for example, collaborative filtering) to another (for example, content based filtering) in the same domain, or across domains" Mediation can be done converting user models and integrating them into a single user model. This entails dealing with syntactic and semantic heterogeneity. [Carmagnola, 11] explain that the translation of user models is managed by a mediator component which is responsible for the following tasks: i) mapping from services to a generic representation and vice versa; ii) providing standard language/interface for the exchange of user model data; iii) maintaining user modeling semantic knowledge for facilitating ad-hoc mapping. Table 1 shows a summary of the techniques used in mediation based approaches

		Approach proposed for mediation between user models				
		Relationships between user models	**Shared User model**	**Multi-agent system**	**Conversational**	**Machine learning**
[Vassileva et al.,99; 01; 03]	iHelp			X		
[Niu et al., 03]				X		
[Trella, et al, 03]	Medea	Manual				
[Cali et al, 04]		DL-based				
[González et al, 05]	Smart User Model [SUM]			X		X
[Berkovsky et al, 06]	Trip@dvisor and PIL		X			
[Van der Sluijs and Houben, 06]	GUC		semi-automatic mappings			
[Stewart et al, 06]		Manual				
[Metha, 07]				X		X
[Carmagnola & Dimitrova, 08]			auto/semiauto map			
[Wang et al, 08]	iCity to CHIP	Lexical Mapping				
[Carmagnola, 09]						
[Cena&Furnari, 12]	UMIS				WS - Dialog Game	

Table 1: Summary of the techniques used in mediation-based approaches

We can mention the authors in table 1 as representatives of this approach. The Generic User model Component (GUC) presented in [Van der Sluijs, 06] proposed to construct combined ontologies for exchanging user models between web-based systems. GUC allows the configuration of a distributed management of mappings between user models. The schema mappings were determined by a human, and the possibility of automatic merging techniques was discussed, but implemented requiring human effort. Carmagnola et al.[Carmagnola, 09] proposed a solution with high flexibility representing user models and providing semantic mapping of the user data from heterogeneous sources. However, to take part in the interoperability process every provider need to comply to a standard format for the exchange, and maintain a sharable user model which includes the fragments of user model as RDF statements.

In summary, we need to provide a semantic representation of a ubiquitous user model that is not static in order to adapt itself to new profile suppliers and consumers, especially in health and fitness domain. We have to deal with semantic heterogeneity with the least intervention and effort of ubiquitous user modeling stakeholders.

3 Schema matching

Bernstein et al. [Bernstein, 11] define:

- ✓ “Schema matching is the problem of generating correspondences between elements of two schemas”.
- ✓ “Schema is a formal structure that represents an engineered artifact, such as a SQL schema, XML schema, entity-relationship diagram, ontology description, interface definition, or form definition.”
- ✓ “A correspondence is a relationship between one or more elements of one schema and one or more elements of the other”

Schema matching techniques are developed to solve problems in many domains and applications such as schema and ontology integration, data integration, data warehousing, knowledge-based and web applications among others.[Rahm, 01; Bernstein, 11].

We are interested in one of the schema matching applications which is schema integration of independently developed schemas. In particular, the integration of ubiquitous user model ontology from heterogeneous and independently developed user model schemas. In fact, future trends in data integration suggest that, as new data sources become available, they can be mapped to a single mediated schema [Bernstein, 11].Most data integration implies therefore matching algorithms to produce correspondences and semantic mappings.

In this section we briefly overview of schema matching and schema integration approaches.

3.1 Schema integration

There is great amount of information available nowadays which is distributed and heterogeneous. “The problem of bringing together heterogeneous and distributed computer systems is known as interoperability problem” [Wache, 01]. Wache et al. explain how in order to achieve information sharing, it is necessary first to provide accessibility to the data, and then process and interpret data. The heterogeneity of data brings structural and semantic heterogeneity; the meaning of the information must be understood to be able to interchange it.

[Noy, 04] presents a brief survey of approaches to semantic integration. It shows how researchers in the databases information-integration and ontology researchers have been addressing the same semantic interoperability issues. As in the user modeling interoperability problem, there are two major approaches to facilitate knowledge integration and sharing: using a shared ontology that provides a common vocabulary for a domain, or finding correspondences between concepts of different schemas or ontologies that puts together information. Both approaches enable data sharing, query answering or web-service composition.

There are two different integration system architectures [Cruz, 05]: i) A central data integration system which has a global schema or ontology with a uniform interface to access information of data sources. ii) A peer to peer data integration where any peer can accept requests of information from any other peer and the information.

Ontologies have and important role in identifying semantic correspondence of information concepts of heterogeneous sources [Wache 01; Cruz, 05; Uschold, 04]. In all integration approaches, ontologies are used to describe the semantics of the data of different information sources in an explicit and machine-understandable conceptualization. [Wache, 01] identify three ways of using ontologies for the integration of several sources: single ontology approaches, multiple ontology approaches and hybrid approaches. In a single ontology approach a global ontology provides the semantic specifications and all source schemas must directly relate to this ontology. All sources must have nearly the same view on a domain with the same level of granularity to participate in the integration. If one source does not meet these requirements, a minimal ontology commitment has to be found [Gruber, 95]. This task is not easy due to domain view or granularity differences.

Each information source is described by its own ontology in the multiple ontology approach. Local ontologies are mapped to each other and no common or minimal ontology is needed. These inter-ontology mapping identifies semantic correspondences between individual terms of different sources which is very difficult to define. Hybrid approaches try to overcome the drawbacks of single and multiple ontology approaches. In this approach, each source defines its own local ontology but it is mapped to a global shared ontology. The main advantage of hybrid approach is that new sources can be easily added without modification of existing mapping [Wache, 01]. This approach is therefore well suited for dynamic environments like the health and fitness environment described in this paper.

3.2 Schema matching approaches

At first, schema matching was performed manually or supported by a graphical interface, but this task is tedious, time consuming and expensive. In the past decade (2000-2010) as the number of web sources increase and databases became more complex, automated support became more necessary.

There are recent surveys [Rahm , 01; Kalfoglou, 03;Shvaiko, 05; Shvaiko ,08; Bernstein, 11; Shvaiko , 13;] and two books [Bellahsene, 11; Euzenat, 07] that reflect the research in the areas of schema and ontology matching which are very closely related.

[Rahm, 01] present a good classification of schema matching approaches which gives us a brief overview of the major approaches of schema matching. An enhanced taxonomy of schema-based matching techniques was presented in [Shvaiko, 05a] based on Rahm et al. previous work. These two classifications of schema matching consider the following orthogonal classification criteria:

- Instance vs. schema. - Matching can be done only with schema information or it can consider instance data.
- Element vs. structure granularity. - Element-level matching is built between individual schema elements. Structural-level matching takes into account the combination and relations between elements.
- Linguistic vs. constraint. - Linguistic technique matching is based on element´s name or description and constrain-based approaches are based on data types, value ranges keys and relationships.
- Matching Cardinality. - Schema matchers can produce 1:1 mappings between schema elements or n:1, 1:n and/or n:m mappings. There may be different match cardinalities also at instance level.
- Using auxiliary information. – Schema matchers can use auxiliary information such as thesauri, input match-mismatch lists or previously computed mappings.
- Individual vs. combinational. - Individual matches use one approach to perform the matching. Individual matchers are usually combined in hybrid matchers (directly combine several matching approaches) or composite matchers (combines results of independent matchers).

[Shvaiko, 05a] presented the classification of schema matching approaches with two new criteria for classification of individual matchers introduced by the authors:

- Heuristic vs. formal. – Heuristic techniques try to guess relations holding between labels or graph structures. Formal techniques have model-theoretic semantics to justify their results.
- Implicit vs. explicit. – Implicit techniques rely on syntax label techniques and explicit techniques exploit semantics of labels.

From the discussion in [Shvaiko, 13] referring to ontology matching challenges and the future trends, and research agenda regarding schema matching from [Bernstein, 11], we extract specific challenges and trends that may be relevant to our work:

- Recent examples of schema and ontology matching solutions use a combination of different similarity measures (hybrid similarity), and schema matching techniques in order to find the commonalities and differences between two concepts.
- An incremental progress of matching systems can be obtained by tuning further similarity measures and finding innovative combinations of matchers. However, breakthroughs can come from completely different settings or systems particularly adapted to specific applications.

4 Ontology and process of concept alignment for ubiquitous user model interoperability

In this section, we present a hybrid approach for schema integration and the global ontology based on Simple Knowledge Organization System (SKOS) [Miles, 07] that provides semantic representation for a flexible ubiquitous user model. We also present an enhanced process of concept alignment that enables interoperability between profile suppliers/consumers and the ubiquitous user model by automatically defining semantic mappings.

We propose to improve the quality of mapping adding an internal structure similarity that considers data type and preferred values constrains. This approach handles syntactic, semantic and schematic heterogeneity of the data, and helps find commonalities and dissimilarities between concepts of different sources.

The user profile structure represented by the ontology is able to evolve over time because the process of concept alignment is capable of establish semantic mappings and determine if new skos:Concepts are to be added to the ontology. So, we try to build a bridge between the two approaches described in section 2 as suggested in [Berkovsky, 09]. We propose a central ubiquitous user model that can adapt itself to changing multi-application environment, and a process of concept alignment that provides articulation between heterogeneous sources without any effort of profile suppliers or consumers. The ontology and process of concept alignment are part of a work-in-progress project of a framework for ubiquitous user models interoperability.

4.1 U2MIO Ontology

We briefly describe Ubiquitous User Modeling Interoperability Ontology (U2MIO) that:

- Provides semantic support for user model overcoming differences between concepts at knowledge level.
- Represents a flexible user profile structure, with domain independency which provides the possibility for the ubiquitous user model to evolve during time.
- Provides representation for new profile suppliers and consumers that take part in the interoperability process without effort of the provider or consumer system.

The ontology reuses SKOS ontology; it can be seen as an aggregation of concept schemes each one representing a profile supplier or consumer, and a central ubiquitous user model concept scheme. Semantic mapping relations are established between each supplier/consumer concept scheme and the ubiquitous user model concept scheme at concept level by the process of concept alignment in order to enable interoperability between user models.

The ubiquitous user modeling ontology was set-up with Facebook, FOAF, and one profile of a specialized web application to monitor person's diet and physical activity of one user. This demands the design of four concept schemes, one for each profile provider and the ubiquitous user model concept scheme. Semantic mapping relations were established with SKOS properties. We used Protégé ontology Editor for the set-up process. A detailed description of U2MIO is given in [Martinez-Villaseñor , 12b].

4.2 Process of concept alignment

The ubiquitous user modeling framework mentioned in section 3, deals with providers' transfer mechanisms and obtains sources documents *(sd)* in XML, JSON or RDF. If the source is new to the system, a corresponding skos:ConceptScheme *(C)* is designed and added to U2MIO. The process of concept alignment automatically determines semantic mappings for each source concept with the best suited concept in the ubiquitous user model concept scheme *(u2m).*

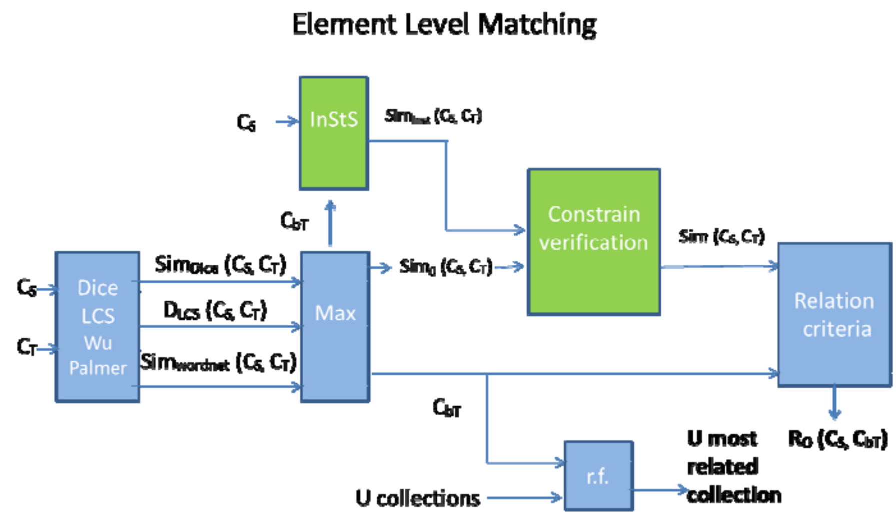

Figure 1: Element level matching phase

New skos:Concepts and skos:Collections can be added to *u2m* resulting of rule based decisions made by the process of alignment.

The process of concept alignment is based in a two-tier matching strategy. First an element level matching step finds a set of concept candidates for alignment for each concept in the source concept scheme (Figure 1). A concept scheme is defined as (C, H_C, V_C) where C is a set of concepts arranged in a subsumption hierarchy H_C. V_C is the set of corresponding concept values. C_s is the set of concept labels extracted from the source document. C_T is the set of concept labels extracted from the target (ubiquitous user model scheme), and C_{bT} is the set of concepts that are best suited for alignment.

The element level matching is performed combining different similarity measures (Sim_{Dice},D_{LCS}, $Sim_{Wordnet}$) and schema matching techniques described in the following subsection. From this step, in which we analyze the word similarity between each concept in the source with all concepts in *u2m*, we find a set of best suited concepts for alignment (or one best suited concept) in the target (*u2m*), and the collection most related with the source document.

Next, the phase looks for structure similarity (Figure 2).

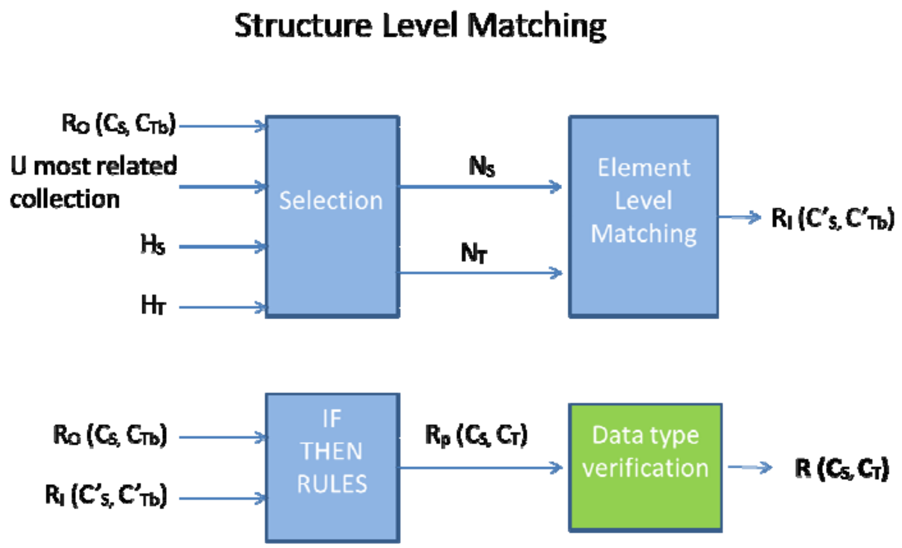

Figure 2: Structure level matching phase.

The goal of the structure level matching step is to disambiguate the meaning of the word analyzing its context, this means analyzing the structure and meaning of the neighbor concepts in the same source document. In this step, the similarity between the neighbor concepts in the source and the neighbors of the best suited concept(s) in the target are calculated.

Given the highest relation R_0 (C_s, C_{Tb})obtained from the element level matching phase and the highest relation $R_1(C'_s, C'_{Tb})$ resulting of the structure level matching similarity calculation, a set of IF THEN rules are applied to determine one-to-one semantic mappings and recommendations of concept and collection additions.

The process of concept alignment is enhanced adding a new similarity measure that considers data type similarity and enumeration constraining facets from the source data. More strict data type verification is done once the process of concept alignment is completed to increase the possibility that concepts considered *exact match* are in fact interchangeable. $R(C_s, C_T)$ are the final semantic mapping relations found between the concepts of the source document and the ubiquitous user model (target). Modifications to the previously presented process are described in the next subsection and the processes are represented with green squares in figure 1 and 2. We described the detail of the process of concept alignment and the similarity equations in [Martinez-Villaseñor, 12].

4.3 An enhanced concept similarity for user model interoperability

The similarity between two concepts is calculated in the element level matching. The main similarity equation combines three types of similarity measures: a) String similarity based in Dice [Dice, 45] b) A simple distance of longest substring c) semantic similarity based on WordNet [Wu, 94].

At this stage the process of concept alignment compares individual concepts and verifies string similarity (Dice), label inclusion (longest substring) and label semantic equivalence (Wu and Palmer). Despite the fact that this combined similarity measure covers three aspects that help the process to identify the best suited concepts for alignment, many possible conflicts as the ones described by [Sosnovsky, 09] can occur when dealing with highly autonomous applications.

In order to address these conflicts, we can exploit the information available in the source schema. We only used the concept labels for the previous similarity calculation and the content model which represents the structure of the elements and their relations are used in the structure level matching to calculate the context similarity.

The XML schema, when available, includes other valuable information like data types, cardinality constraints, and other constraining facets like pattern or enumeration. Thuy [Thuy, 12] proposed an interesting hybrid similarity measure for XML data integration and transformation that includes new metrics to compute the data type and cardinality constraint similarities.

We consider that a comparison of the internal structure concepts can help in making sense of the meaning in case of conflict occurrences. We use equation 1 to determine the internal structure similarity of two concepts. It considers a weighted similarity that takes into account the data type similarity and the compatibility between enumerated values when available.

$$sim_{inSt}(c_s, c_{tb}) = \alpha_1(vcf(c_s, c_{tb})) + \alpha_2(sim_{type}(c_s, c_{tb})) \forall c_s \in C_S, \forall c_{tb} \in C_T \quad (1)$$

In equation 1, α_1 and α_2 are weights given to determine the relevance of each factor and the sum of both values $\in$ [0,1]. In this case $\alpha_1 =$ .8 and $\alpha_2 =$.2 which means that the fact that two concepts have enumerated values in common is more relevant than the type compatibility at this stage of the alignment process. Equation 2 determines the value constrain factor *vcf* that verifies if the enumerated or vocabulary values given for a concept in the source, when available, appear also as constrain in the target side by the best suited concept for alignment resulting from the similarity of the element level matching process.

$$vfc(c_s, c_{tb}) = \frac{nv_{st}}{nv_s} \forall c_s \in C_S, \forall c_{tb} \in C_T \quad (2)$$

In equation 2: nv_{st} is the number of enumerated values in the source concept c_s that can be also found in the best suited concept in the target c_{tb} and nv_s is the number of enumerated values in the source. Equation 2 does not apply when nv_s =0.

The data type similarity sim_{type} is extracted from the data type compatibility table presented in [Thuy, 12].
The process of constrain verification pictured in figure 1 in a green square rectifies the similarity value of a given best suited concept for alignment comparing the internal structure similarity (sim_{inSt}) with the combined similarity (sim_0). The following IF THEN rule is used for the comparison:

1. if $sim_{inSt}(c_s, c_{tb}) > sim_0(c_s, c_{tb})$ then $sim_0(c_s, c_{tb}) = sim_{inSt}(c_s, c_{tb})$
2. else
3. if $sim_0(c_s, c_{tb}) > .9 and (vfc(c_s, c_{tb}) < .5 or sim_{type}(c_s, c_{tb}) < .3)$
4. the c_{tb} is discarded as best suited concept for alignment

The condition of $vfc(c_s, c_{tb}) < .5$ only applies if a value constrain is given in the target concept.

Final data type verification is made after the structure level matching process is concluded. If an exact match is found between two concepts and sim_{type} <1 then the final relation is adjusted to *closeMatch* instead of *exactMatch* to increase the feasibility of interchange when an *exactMatch* is found.

5 Experimental Results

In our experiments, we illustrate how reusing and sharing information in multiple user models can be done with the mediation of our ubiquitous user model and the process of concept alignment. In this scenario, the user usually has to endure the pain of repeatedly setting new web applications explicitly capturing basic demographic data and downloading again the valuable data of many training sessions, captured with wearable devices that are not always compatible with the new web application.

We choose to align a new profile provider to the ubiquitous user model that is an especially tough given that all concept labels can be confusing even to a human expert. The source document refers to a fitness profile extracted with the profile.get method from the Fatsecret platform API[1]. It includes several weight values and labels like *weight_measure* and *height_measure* that are frequently used in other similar application with a different sense. This XML source document contains seven concepts.

Once the process aligns this concept scheme, information from this application is reused by an even tougher profile consumer: the weight measurement complex type of Microsoft HealthVault [2]. This weight document has twenty concepts and must be aligned with the ubiquitous user model for mediation also.

Two experiments are done in order to determine the influence of the internal structure similarity in the efficiency and effectiveness of the process of concept alignment.

---

[1] http://platform.fatsecret.com/api/Default.aspx?screen=rapiref&method=profile.get

[2] http://developer.healthvault.com/pages/types/viewsamplexml.aspx?name=Weight%20Measurement&id=3d34d87e-7fc1-4153-800f-f56592cb0d17

5.1 Process of Concept Alignment Evaluation Criteria

In order to measure the efficiency and effectiveness of the matching/mapping systems, different metrics have been proposed in the literature [Bellahsene, 11].

In this work, we focus the evaluation of the process of concept alignment in:

- The human effort required by the mapping designer to verify the correctness of the mappings, which is quantified with the metric *overall* [Bellahsene, 11] and partially measures the efficiency of our process.
- The quality of the generated mappings quantifying the proximity of the results generated by the process of concept alignment to those expected with four known metrics: *precision*, *recall*, *f-measure* and *fall-out* [Euzenat, 07].With these metrics we partially measure the effectiveness of our process.

These metrics are based on the notions of *true positives* (TP), *false positives* (FP), *true negatives* (TN) and *false negatives* (FN).

A human expert provided a list of expected matches for the proof of concept example and evaluated the outcomes, deciding if the semantic mapping relations found were correct and recommendations make sense. Exact match relations correctly found by the process, and good recommendations for concept or collection addition were considered as *TP*. Wrong exact matches were listed as *FP*. When a relevant exact match was not found by the process, a concept was improperly discarded or a wrong recommendation was made, it was registered as *FN*. Properly discarded concepts were recorded as *TN*.

5.2 Results

The results in terms of previously mentioned criteria, metrics and conditions are presented in tables 2-7 and figures 3 and 4. Table 2 shows the resulting confusion matrix of the process of concept alignment between the concept scheme modeled from Fatsecret XML document described for profile.get provider's method, and the ubiquitous user model concept scheme.

		Expected matches	
		positive	negative
Process of concept alignment outcome	positive	TP = 4	FP = 2
	negative	FN = 1	TN = 0

Table 2: Confusion matrix without internal structure analysis when aligning FatSecret

This experiment is done with the previous process of concept alignment without the internal structure analysis enhancement. One *FN* and one *FP* value result from the confusing naming of the concepts and not analyzing the enumerated values. The other *FP* value results from a longest common substring that needs to be improved (it does not consider word token combinations so the *lcs* string between *weight_goal* and *goal_weight* is *weight*).Table 3 shows the results of the same Fatsecret concept scheme alignment using the enhanced process of concept alignment with the internal structure analysis. The enhanced process corrected one *FN* and one *FP*.

		Expected matches	
		positive	negative
Process of concept alignment outcome	positive	TP = 5	FP = 1
	negative	FN = 0	TN = 1

Table 3: Confusion matrix with internal structure analysis when aligning FatSecret

Table 4 demonstrates the influence of the internal structure analysis of the enhanced process of concept alignment. We can observe that all of the metrics considered show an improvement in the quality of generated mappings and efficiency of the process. Figure 4 presents a graphic of these results.

Measure	**Metric**	**Results**	
		With internal structure analysis	**Without internal structure analysis**
Quality of generated mappings (Effectiveness)	*Precision*	83%	67%
	Recall	100%	80%
	F-measure	91%	73%
	Fall-out	50%	100%
Human effort (Efficiency)	*Overall*	80%	40%

Table 4: Influence of internal structure analysis in efficiency and effectiveness results (FatSecret)

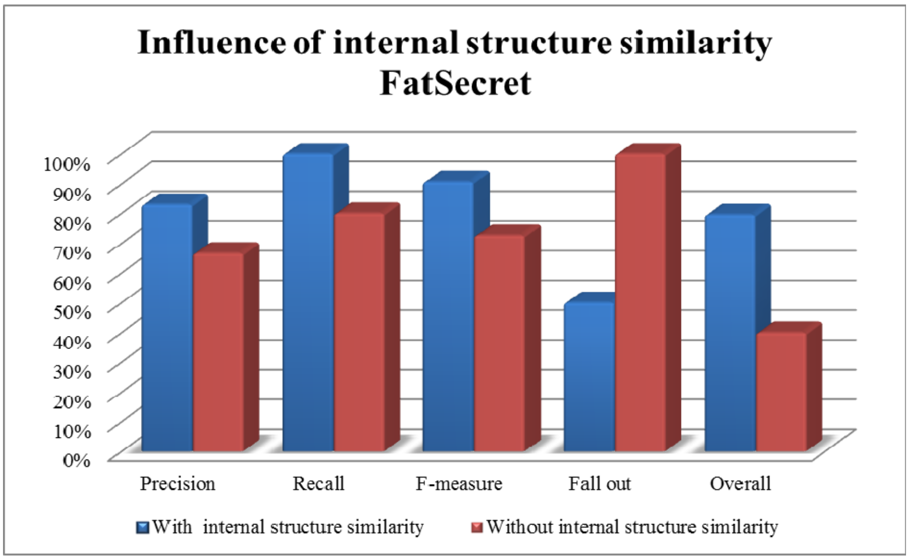

Figure 3: Influence of internal structure analysis FatSecret

Similarly, the second experiment was performed with the consumer´s concept scheme of Microsoft HealthVault weight measure.

The concept labels are not easily detected as meaningful. For example the date is decomposed in single integers for year, month, day, minute, second and f. The concept labels are just the first letter of the word. Even with the ancestor label (date and time), it is difficult to establish the relation with the date the weight measure was taken. Complex types are not easily handled automatically. It also includes ten concepts related with strict meaning for the provider.

Table 5 shows the results of the process of concept alignment of Microsoft Health Vault weight measure concept scheme without the internal structure analysis and table 6 presents the results taking this analysis into account.

		Expected matches	
		positive	negative
Process of concept alignment outcome	positive	TP = 8	FP = 5
	negative	FN = 7	TN = 0

Table 5: Confusion matrix without internal structure analysis when aligning MS HealthVault

Although many difficulties were found due to naming confusions, complex types and not relevant concepts, we can also observe in table 7 and figure 4, that the effectiveness and efficiency were improved by the influence of the internal structure analysis.

		Expected matches	
		positive	negative
Process of concept alignment outcome	positive	TP = 11	FP = 2
	negative	FN = 6	TN = 1

Table 6: Confusion matrix with internal structure analysis when aligning MS HealthVault

Measure	**Metric**	**Results**	
		With internal structure similarity	**Without internal structure similarity**
Quality of generated mappings (Effectiveness)	*Precision*	85%	62%
	Recall	65%	53%
	F-measure	73%	57%
	Fall-out	67%	100%
Human effort (Efficiency)	*Overall*	52%	20%

Table 7: Influence of internal structure analysis in efficiency and effectiveness results (MS HealthVault)

Three FP values were corrected by the enhanced process of concept alignment in this case.

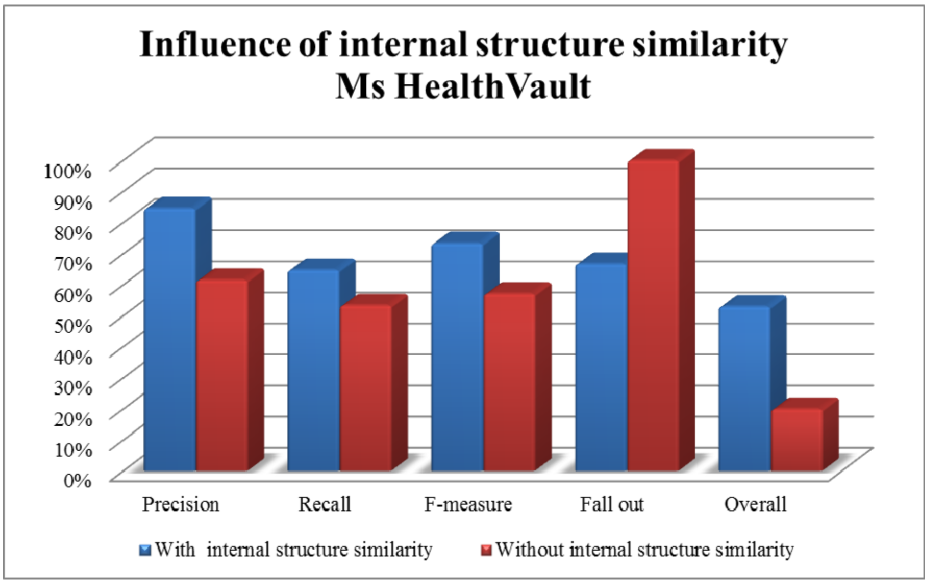

Figure 4: Influence of internal structure analysis MS HealthVault

6 Conclusions and Future Work

We present an enhanced process of concept alignment that considers internal structure of source data for the schema integration with a ubiquitous user model. Despite the fact that previous similarity measures cover lexical, structural, and semantic measures, they were not enough to solve conflicts caused by confusing names, complex types and granularity differences.

Two modifications were made to the previous process of concept alignment: A new combined internal structure similarity measure was added and implemented in the element level matching process; the addition of a final data type verification process to the structure level matching process. We are measuring if two concepts have enumerated or preferred values in common, and the similarity between data types. The first comparison can help disambiguating the meaning in case of occurrences of the mentioned conflicts. The second addition increases the feasibility that data are really interchangeable when an exact match is found. Constrain and data type verification done in the internal structure analysis proved to be useful when dealing with common conflicts between concepts.

Our experimentation shows that, in both cases, the influence of the analysis of data type and enumerated value constrain, combined similarity, and the final data verification, improve the quality of the mappings and decrease the human effort that must been made to correct the process errors. In future work, more exhaustive evaluation must be done.

The XML schema often includes other valuable information like cardinality constraints, and other constraining facets that can be taken into account in the future to improve the quality of the mappings found by the process of concept alignment.

References

[Berkovsky, 06] Berkovsky,S.: Decentralized Mediation of User Models for a Better Personalization. In Adaptive Hypermedia and Adaptive Web-Based Systems, ed. Vincent Wade, Helen Ashman, and Barry Smyth, 4018:404–408. Springer Berlin / Heidelberg. 2006. http://dx.doi.org/10.1007/11768012_59.

[Berkovsky, 09] Berkovsky, S., Heckmann,D. and Kuflik,T.: Addressing Challenges of Ubiquitous User Modeling :Between Mediation and Semantic Integration. Advances in Ubiquitous User Modelling 5830: 1–19. 2009. doi:10.1007/978-3-642-05039-8_1.

[Bravo, 06] Bravo, J.,Alamán,X.,Riesgo,T.: Ubiquitous Computing and Ambient Intelligence: New Challenges for Computing, Journal of Universal Computer Science, vol.21 (3),233-235,2006. http://www.jucs.org/jucs_12_3/ubiquitous_computing_and_ambient.

[Bernstein, 11] Bernstein, P. A, Madhavan, J. and Rahm, E.: Generic Schema Matching, Ten Years Later. Proceedings of the VLDB Endowment 4 (11): 695–701. 2011

[Bellahsene, 11] Bellahsene, Z., Bonifati,A., Duchateau, F. and Yannis Velegrakis: On Evaluating Schema Matching and Mapping. In Schema Matching and Mapping, ed. Zohra Bellahsene, Angela Bonifati, and Erhard Rahm, 253–291. Springer, 2011.

[Cali, 04] Cali, A., Calvanese, D. Colucci,S. Di,T. and Francesco,N.: A Description Logic Based Approach for Matching User Profiles. In Proceedings of the International Workshop on Description Logics (DL 2004), 1–10. 2004.

[Carmagnola, 08] Carmagnola, F., and Dimitrova,V.: An Evidence-Based Approach to Handle Semantic Heterogeneity in Interoperable Distributed User Models. In Adaptive Hypermedia and Adaptive Web-Based Systems, 5th International Conference, AH 2008, ed. Wolfgang Nejdl, Judy Kay, Pearl Pu, and Eelco Herder, 73–82. 2008. Hannover, Germany: Springer. doi:isbn:978-3-540-70984-8.

[Carmagnola, 09] Carmagnola, F.: Handling Semantic Heterogeneity in Interoperable Distributed User Models. Advances in Ubiquitous User Modelling 5830: 20–36. 2009. doi:10.1007/978-3-642-05039-8_2. http://www.springerlink.com/content/g227v147n06rx619/.

[Carmagnola, 11] Carmagnola, F., Cena, F. and Gena,C.: User Model Interoperability: a Survey. User Modeling and User-Adapted Interaction 21 (3) (February 18): 285–331. 2011. doi:10.1007/s11257-011-9097-5. http://www.springerlink.com/index/10.1007/s11257-011-9097-5.

[Cena, 09] Cena, F., and Furnari, R. : A Model for Feature-Based User Model Interoperability on the Web. In Advances in Ubiquitous User Modelling, ed. Tsvi et al. Kuflik, 5830:37–54. Springer Berlin / Heidelberg. 2009. http://dx.doi.org/10.1007/978-3-642-05039-8_3.

[Cruz, 05] Cruz, I. F, and Xiao, H.: The Role of Ontologies in Data Integration. Journal of Engineering Intelligent Systems 13: 2005. 245–252.2005.

[Dice, 45] Dice, L.R. 1945. Measures of the Amount of Ecologic Association between Species. Ecology 26.

[Dolog, 05] Dolog, P, and M Schäfer.: A Framework for Browsing, Manipulating and Maintaining Interoperable Learner Profiles. In Proceedings of UM2005 10th International Conference on User Modeling.2005.

[Euzenat 07] Euzenat, J., and Shvaiko, P.: Ontology Matching. Springer-Verlag, Berlin Heidelberg.333 pp. 2007. ISBN: 978-3-540-49611-3.

[González, 05] González, G., López, B. and J.: A Multi-agent Smart User Model for Cross-domain Recommender Systems. In Proceedings of Beyond Personalization: The Next Stage of Recommender Systems Research, International Conference on Intelligent User Interfaces IUI'05. San Diego, California, USA. 2005

[Gruber, 95] Gruber, T. R. 1995. : Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies - Special issue: the role of formal ontology in the information technology. 43 (5-6): 907–928.1995. doi:10.1006/ijhc.1995.1081. http://dx.doi.org/10.1006/ijhc.1995.1081.

[Heckmann, 05] Heckmann, D.: Ubiquitous User Modeling". Ph.D Dissertation ,Department of Computer Science, Saarland University Germany, Saarland, Germany. 2005.

[Kay, 03] Kay, J., Kummerfeld, R. J. and Lauder,P.: Managing Private User Models and Shared Personas. In Proceedings of the Workshop on UM for Ubiquitous Computing, 22–26. Pittsburg,PA,USA. 2003.

[Kalfoglou, 03] Kalfoglou, Y., and Schorlemmer, M.: Ontology Mapping: The State of the Art. The Knowledge Engineering Review 18 (1): 1–31. 2003. doi:10.1017/S0269888903000651. http://dx.doi.org/10.1017/S0269888903000651.

[Martinez-Villaseñor, 12] Martinez-Villaseñor, M.L., and Gonzalez-Mendoza,M, and Hernandez-Gress,N.: Towards a Ubiquitous User Model for Profile Sharing and Reuse. Sensors 12 (10): 13249–13283. 2012. doi:10.3390/s121013249. http://www.mdpi.com/1424-8220/12/10/13249.

[Martinez-Villaseñor, 12a] Martinez-Villaseñor, M.L., and Gonzalez-Mendoza,M.: Process of Concept Alignment for Interoperability Between Heterogeneous Sources, 11th Mexican International Conference on Artificial Intelligence, MICAI 2012, San Luis Potosí, México,October 27–November 4,LNAI 7669, 310-319. 2012a.

[Martinez-Villaseñor, 12b] Martinez-Villaseñor, M.L., and Gonzalez-Mendoza,M.: Towards an Ontology for Ubiquitous User Modeling Interoperability." In KEOD 2012 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, 4 - 7 October, ed. J. Filipe and J. L. G. Dietz, 239–244. Barcelona, Spain: SciTePress. doi:isbn:978-989-8565-30-3. 2012b.

[Mehta, 05] Mehta, B., Niederee,C., Stewart,A., Degemmis,M., Lops,P. and Semeraro,G.: Ontologically-Enriched Unified User Modeling for Cross-System Personalization. In User Modeling 2005, ed. L. Ardissono, P. Brna, and A. Mitrovic, 3538:119–123. Springer Berlin / Heidelberg. 2005. http://dx.doi.org/10.1007/11527886_16.

[Mehta, 07] Mehta, B.: Learning from What Others Know: Privacy Preserving Cross System Personalization. In Proceedings of the 11th International Conference on User Modeling, 57–66. Berlin, Heidelberg: Springer-Verlag. 2007. doi:10.1007/978-3-540-73078-1_9. http://dx.doi.org/10.1007/978-3-540-73078-1_9.

[Miles, 07] Miles, A., and Pérez-Agüera, J. R.: "SKOS: Simple Knowledge Organization for the Web. Cataloging & Classification Quarterly 43 (3-4): 69–83. 2007. doi:doi:10.1300/J104v43n03_04. http://www.tandfonline.com/doi/abs/10.1300/J104v43n03_04.

[Niu, 03] Niu, X., Mccalla,G. and Vassileva,J.: Purpose-based User Modelling in a Multi-agent Portfolio Management System.2003. In Proceedings of UM'03 Proceedings of the 9th international conference on User modeling,398-402, Springer-Verlag Berlin, Heidelberg, 2003.

[Noy, 04] Noy, N. F.: Semantic Integration: a Survey of Ontology-based Approaches.” SIGMOD Record 33 (4),: Stanford University, Stanford, CA, 65–70. December 2004. doi:10.1145/1041410.1041421. http://doi.acm.org/10.1145/1041410.1041421.

[Rahm, 01] Rahm, E., and Bernstein, P.: A Survey of Approaches to Automatic Schema Matching. The VLDB Journal 10 (4): 334–350.2001 http://dx.doi.org/10.1007/s007780100057.

[Razmerita, 03] Razmerita, L., Angehrn, A. and Maedche, E.: Ontology-based User Modeling for Knowledge Management Systems. In Proceedings of the Ninth International Conference of User Modeling 2702: 213–217. 2003.

[Rui, 10] Rui, J., Rodrigues, H., Otero, N.: Ambient Intelligence: Beyond the Inspiring Vision, Journal of Universal Computer Science vol. 16 (12), 1480-1499, June, 2010. http://www.jucs.org/jucs_16_12/ambient_intelligence_beyond_the.

[Shvaiko, 05] Shvaiko, P., and Euzenat, J.: A Survey of Schema-Based Matching Approaches. Ed. Stefano Spaccapietra. Journal on Data Semantics IV 3730: 146–171, 2005. doi:10.1007/11603412_5. http://dx.doi.org/10.1007/11603412_5.

[Shvaiko, 05a] Shvaiko, P.,: A Classification of Schema-based Matching Approaches. Journal on Data Semnatics (4).146-171, 2005.

[Shvaiko, 06] Shvaiko, P., Giunchiglia, F., Schorlemmer, M., McNeill, F., Bundy A., Marches, M., Yatskevich, M. et al.: OpenKnowledge Deliverable 3.1.: Dynamic Ontology Matching: A Survey. Department of Information Engineering and Computer Science, DIT-06-046, 27 July 2006.

[Shvaiko, 08] Shvaiko, P. and Euzenat, J.: Ten Challenges for Ontology Matching. In On the Move to Meaningful Internet Systems: OTM 2008, ed. Robert Meersman and Zahir Tari, Springer Berlin / Heidelberg, 2008. 5332:1164–1182. http://dx.doi.org/10.1007/978-3-540-88873-4_18.

[Shvaiko, 13] Shvaiko, P. and Euzenat, J.: Ontology Matching: State of the Art and Future Challenges. Knowledge and Data Engineering, IEEE Transactions On 25 (1), TrasLab, Inf. Trentina, Trento, Italy: 158–176. 2013. doi:10.1109/TKDE.2011.253.

[Sosnovsky, 09] Sosnovsky, S., Brusilovsky, P. Yudelson, M., Mitrovic, A. Mathews, M. and Kumar, A., Semantic Integration of Adaptive Educational Systems. In Advances in Ubiquitous User Modelling, ed. A Kuflik, T., Berkovsky, S., Carmagnola, F., Heckmann D., & Krüger, 134–158. LNCS 5830. Berlin Heidelberg: Springer-Verlag, Berlin Heidelberg. 2009.

[Stewart, 06] Stewart, C., Cristea, A., Celik, I. and Ashman, H.: Interoperability Between AEH User Models. In Proceedings of the Joint International Workshop on Adaptivity, Personalization & the Semantic Web, 21–30. Odense, Denmark, 2006.: ACM. doi:10.1145/1149933.1149937. http://doi.acm.org/10.1145/1149933.1149937.

[Thuy, 12] Thuy, P.: Hybrid Similarity Measure for XML Data Integration and Transformation”. Ph.D. Dissertation. Department of Computer Engineering, Kyung Hee University, Seoul, Korea. 2012.

[Trella, 03] Trella, M., Cornejo, R., Guzman, E. and Bueno, D. 2003. An Educational Component Based Framework for Web ITS Development. In Proceedings of the 2003 International Conference on Web Engineering ICEW 2722: Springer-Verlag, Heidelberg, 134–143,2003.

[Uschold, 04] Uschold, M., and Gruninger, M.: Ontologies and Semantics for Seamless Connectivity. SIGMOD Rec. 33 (4): 58–64. 2004. doi:10.1145/1041410.1041420. http://doi.acm.org/10.1145/1041410.1041420.

[Van der Sluijs , 06] Van der Sluijs, K., and Houben.G.J.: A Generic Component for Exchanging User Models Between Web-based Systems. International Journal of Continuing Engineering Education and Life Long Learning 16 (1/2): 64–76.2006. http://wwwis.win.tue.nl/~houben/respub/ijceell2006.

[Vassileva, 01] Vassileva, J. : Distributed User Modelling for Universal Information Access. In Proceedings of the 9th International Conference on Human-Computer Interaction, Ed. Lawrence Erlbaum, 122–126. New Orleans, USA. 2001.

[Viviani, 10] Viviani, M., Bennani, N. and Egyed-Zsigmond, E.: A Survey on User Modeling in Multi-application Environments. In Proceedings of the 2010 Third International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies and Services, Nice, France. 111–116. 2010. Ieee.doi:10.1109/CENTRIC.2010.30.http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5600339.

[Wache, 01] Wache, H, Vögele, T. ,Visser, U Stuckenschmidt, H. ,Schuster, G., Neumann, H. and Hübner, S.: Ontology-based Integration of Information - a Survey of Existing Approaches. In IJCAI--01 Workshop: Ontologies and Information Sharing, ed. H Stuckenschmidt, 108–117, 2001. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.8073.

[Wang, 08] Wang, Y., Cena, F., Carmagnola, F,. Cortassa, O., Gena,C., Stash, N. and Aroyo, L.: RSS-Based Interoperability for User Adaptive Systems. In Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Ed. Wolfgang et al. Nejdl, Berlin, Heidelberg: Springer-Verlag, 353–356. 2008. doi:10.1007/978-3-540-70987-9_52. http://dx.doi.org/10.1007/978-3-540-70987-9_52.

[WHO Regional Office for Europe, 07] WHO Regional Office for Europe: The Challenge of Obesity in the WHO European Region and the Strategies for Response. 2007. Available at http://www.euro.who.int/document/E90711.pdf.Accesed on February 10th, 2012.

[Wu, 94] Wu, Z., and Palmer, M. :Verb Semantics and Lexical Selection. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, NM, USA. 133–138,1994.