On Constructing, Grouping and Using Topical Ontology for Semantic Matching

Yan Tang, Peter De Baer, Gang Zhao, and Robert Meersman

VUB STARLab
10G731, Vrije Universiteit Brussel
Pleinlaan 2, 1050 Elesene
Brussels, Belgium
{yan.tang, pdebaer, robert.meersman}@vub.ac.be,
gang.zhao@intelartes.com

Abstract. An ontology topic is used to group concepts from different contexts (or even from different domain ontologies). This paper presents a pattern-driven modeling methodology for constructing and grouping topics in an ontology (PAD-ON methodology), which is used for matching similarities between competences in the human resource management (HRM) domain. The methodology is supported by a tool called PAD-ON. This paper demonstrates our recent achievement in the work from the EC Prolix project. The paper approach is applied to the training processes at British Telecom as the test bed.

1 Introduction and Motivation

In the EC Prolix project[1], we need to calculate competency gaps between a learning module (such as a learning material or a course), a person's profile (such as his Curriculum Vitae), and the descriptions of a job in the human resource management (HRM) domain. An HRM ontology is required for calculating this kind of gaps and providing semantic reasoning.

We use the concept of topical ontology and ontological topics [14] as they provide freely combined, easily manageable concept sets, which can be from different contexts and even from different ontologies.

In this paper, we will illustrate a pattern-driven topical ontology creation methodology (PAD-ON), its supported tool PAD-ON suite and how to use the topical ontology in the simple semantic matching algorithm. It is applied to a use case from British Telecom in Amsterdam[2] – finding the conceptual similarities between the learning materials and assessment materials. The paper is organized as follows. Chapter 2 is the paper background. We present PAD-ON in chapter 3. Chapter 4

[1] The EC Prolix (FP6-IST-027905, Process-Oriented Learning and Information Exchange, http://www.prolixproject.org/) is project co-funded by the European Commission under the Sixth Framework Program. It is to align learning with business processes in order to enable organisations to faster improve the competencies of their employees according to continuous changes of business requirements.

[2] http://www.bt.com/

R. Meersman, P. Herrero, and T. Dillon (Eds.): OTM 2009 Workshops, LNCS 5872, pp. 816–825, 2009.

discusses how to use the resultant ontology in the semantic matching algorithm. Chapter 5 is the result of the paper. The paper related work and the comparison between our work and the others' are discussed in Chapter 6. We conclude and illustrate our future work in chapter 7.

2 Background

2.1 Ontology and DOGMA

An ontology is a specification of a conceptualization [7]. The concepts, relationships, and other distinctions relevant for modelling a domain are defined and specified.

In the DOGMA (Developing Ontology-Grounded Methods and Applications, [9, 11,13]) framework, one constructs (or converts) ontologies by the *double articulation* principle into two layers: 1) the lexon base layer that contains a vocabulary of simple facts called *lexons*, and 2) the *commitment* layer that formally defines rules and constraints by which an application (or "agent") may make use of these lexons.

A lexon is a quintuple $\langle \gamma, t_1, r_1, r_2, t_2 \rangle$, where γ is a context identifier, within which the terms t_1, t_2 are defined. The role r_1 and co-role r_2 refer to the relations that the concepts share with respect to one another.

For example, a lexon $\langle \gamma, driver'slicense, is\ issued\ to, has, driver \rangle$ a fact that "a driver's license is issued to a driver", and "a driver has a driver's license".

A *commitment* contains a set of rules in a given syntax, and describes a particular application view of reality, such as the use by the application of the (meta-) lexons in the lexon base, e.g. we may apply the *uniqueness* constraint on the lexon in order to have has a constraint - "*one* driver's license is issued to *at most one* driver".

Ontologies modeled in DOGMA can be further implemented in an ontology language, such as OWL[3] and RDF(S)[4].

The specification (as defined by Gruber [7]) takes the form of the definitions of representational vocabulary (such as lexons and commitments in DOGMA), which provide meanings for the vocabulary and formal constraints on its coherent use.

2.2 Topical Ontology

A topical ontology is specified into five blocks of specifications (Fig. 1). They are the basis, assertions, topics, applications and instances described as below.

- The specification of **Basis** identifies the conceptual objects and their typological relationship. In DOGMA, the basis is a lexon set that only contains lexons with "is-a" (or "part-of", "subtype", "subclass" etc.) relationships, for instance, lexon $\langle \gamma, employee, is\ a, is, person \rangle$. This kind of lexons can be represented graphically as a hierarchical, acyclic tree.
- The **Assertions** asserts how the conceptual objects interact with each other. It contains the lexons that have "free" relations or playing "free" roles with each other, e.g. lexon $\langle \gamma, trainer, teach, is\ taught\ by, course \rangle$. This kind of lexons can be represented graphically as a non-directed network.

[3] OWL Web Ontology Language: http://www.w3.org/TR/owl-features/
[4] RDF(S) Resource Description Framework (Schema): http://www.w3.org/TR/rdf-schema/

- The **Topics** groups the relationships and assertions into a topical map of larger reusable units. We consider a topic as a collection of lexons. Although the theoretical basis for topics is different from that of a context, a topic can sometimes play the role of a context.
- The **Applications** select, constrain and contextualize topics, relationships and assertions to formulate abstracts about business data and processes in view of specific application semantics. It is a set of ontological commitments that contain constraints and axioms.
- The **Instance** lists concrete references or values about conceptual objects, relationships and assertions. For example, an instance of "skill level" in the lexon $\langle \gamma, skill, has, is\ of, skill\ level \rangle$ can be "good".

Note that a lexon in the block of Instance, Application and Topics needs to be defined in either the block of Basis or the block of Assertions, or both.

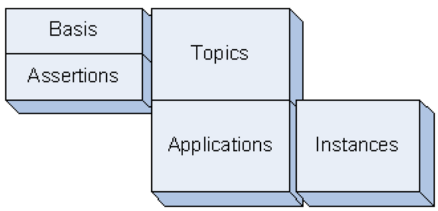

Fig. 1. Ontology building blocks

3 Pattern-Driven Topical Ontology Modeling Methodology (PAD-ON)

This methodology contains three steps: 1) create pattern(s); 2) create ontological basis (subtypes of each element in the patterns); 3) create ontological assertions; 4) group lexons in the ontological basis and the ontological assertions into topics; 5) model ontological commitments for the applications and populate the instances of some concepts.

The design pattern in this paper is an extension to the design pattern discussed in [14], which consists of five elementary elements: Participant, Action/Process, Object, Attribute and State (Fig. 2).

On step 1, our task is to define the patterns as illustrated above. We can use different resources, such as the Organization for Economic Co-Operation and Development (OECD[5]), and the O*NET[6], to define the patterns based on Fig.2.

For instance, we extend the pattern in Fig.2 with OECD key competence framework, the conceptualization of which focuses on the person and his action, the manner and means. It stresses the interaction between individuals at work. We refine Participant into Person. The State of the action is further detailed into the *Instrument* and *Manner*. Accordingly, we have a pattern as illustrated in Fig.3.

[5] http://www.oecd.org
[6] http://online.onetcenter.org/

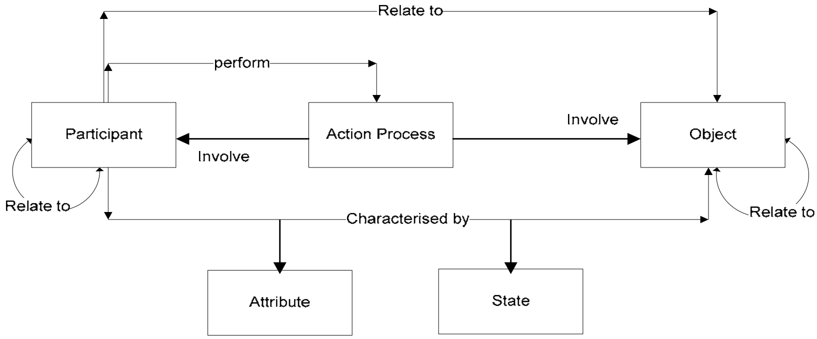

Fig. 2. Pattern used for HRM ontology

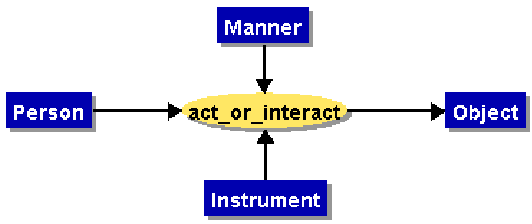

Fig. 3. A pattern in Conceptual Graph [12]: Persons act or interact on objects by instruments in manners

The elements in Fig.3 are specified as below.

- **Person/Actor:** an initiator of an event or performer of an action, it is a Participant or Person
- **Act/Interact:** an action or even, in which a Participant or Person is lined to another Participant or Person.
- **Object:** an affected or resultant entity in an action, event or process. It has no volition and is inanimate.
- **Instrument:** a means with which an Actor causes an event or performs an action. It pertains to the State of an action, process or event.
- **Manner:** a state or form of an action, process or event. It pertains to the State of an action, process or event.

On step 2, we define the subtypes of each element in the above pattern. For instance, the subtypes of "Person" (Fig.3.) are defined in Table 1.

Table 1. Lexon table that contains the subtypes of „Person" (Basis level)

Head Term	***Role***	***co-role***	***Tail Term***
Customer	is a	super type of	Person
Employee	is a	super type of	Person
junior employee	is a	super type of	employee
senior employee	is a	super type of	employee
…	…	…	…

On step 3, we create lexons that does not contain hierarchical relationships as roles as the ontological assertions (Table 2).

Table 2. Lexon table that contains lexons at the Assertion level

Head Term	*Role*	*Co-Role*	*Tail Term*
Person	differentiate	is differentiated by	performance level
performance	has	is of	performance level
Person	make	is made by	Mistake
Person	accept	is accepted by	Mistake
Person	learn from	is learned from	Mistake
Person	give	is given by	Feedback
…	…	…	…

On step 4, we create ontological topics. The topics are considered as freely combined contexts, each of which accumulates lexons that are described at the basis level and assertion level into a set. There are two types of topics in the BT use case – assessment capacities and learning materials. There are in total 10 capacities: Trustworthy, Coaching for performance, Helpful, Bottom Line, Inspiring, Drive for results, Straightforward, Customer connected, Heart, Professional/ technical. For instance, "Coaching for performance" contain the lexons as shown in the table below.

Table 3. Lexon table of topic „Coaching for performance"

Head Term	*Role*	*co-role*	*Tail Term*
person	observe	is observed by	agenda
person	ally with	is allied with	person
person	negotiate with	is negotiated with	person
person	Support	is supported by	person
person	Replace	is replaced by	person
person	Give	is given by	promise
…	…	…	…

The learning materials are linked to BT learning skills, which can be categorized into 4 groups: 1) Process skills, such as knowledge of fault handling, 2) Soft skills, such as ability to cooperate, 3) Technical skills, such as voice-T1 and DSL-T4, 4) Administrative and other skills, such as the ability of handle the calendar and the ability of making customer report. For instance, learning material "Senior SAT" correspond to the ontological topic "Senior SAT" as shown in Table 4.

Table 4. Lexon table of topic „Senior SAT"

Head term	*Role*	*co-role*	*Tail term*
person	use	is used by	Information
person	identify	is identified by	Information

Table 4. (*continued*)

person	locate	is located by	Information
person	access	is accessed by	Information
TS page	is a	super type of	Information
sheet information	is a	super type of	Information
network drawing	is a	super type of	Information
…	…	…	…

On step 5, we model ontological commitments that make use of the lexons created in the previous steps. Note that these commitment models need to be further implemented and integrated into the real applications.

4 Ontological Topics for Semantic Matching

Currently, we use a tool called DMatch, which contains lexical matching algorithms, such as Levenshtein distance [8], and set comparison supported by thesaurus, such as WordNet [5].

The discussion on the matching algorithms and evaluation of different matching algorithms is out of the scope of this paper. The principles of using topical ontology for semantic matching are the issues that need to be discussed here.

With topical ontology, the matching can happen between two topics. For instance, we can find the conceptual similarity between assessment capacity "Coaching for performance" (topic "Coaching for performance", which contains the lexons in Table 3) and learning material "Senior SAT" (topic "Senior SAT", which contains the lexons in Table 4).

We take one lexon from each topic and compare their lexon terms. Below is a list of lexons where lexon $l \in L$ and topic $\zeta \in Z$

$$l_1 = \langle \gamma, person, make, is\ made\ by, mistake \rangle \in \zeta_1$$
$$l_2 = \langle \gamma, person, observe, is\ observed\ by, agenda \rangle \in \zeta_2$$
$$l_3 = \langle \gamma, manager, use, is\ used\ by, schedule \rangle \in \zeta_3$$
$$l_4 = \langle \gamma, senior\ employee, use, is\ used\ by, agenda \rangle \in \zeta_4$$
$$l_5 = \langle \gamma, junior\ employee, work\ with, is\ worked\ on\ by, SAT \rangle \in \zeta_5$$

l_1 is similar to l_2 because they both contain the lexon term "person". l_3 is similar to l_2 as the term "schedule" from l_3 and the term "agenda" from l_2 are in the same SynSet. l_4 and l_5 are similar because both the term "senior employee" from l_4 and the term "junior employee" from l_5 contain the string "employee".

The roles are used for the comparison as well, but in a weighted way. For instance, we can give the weights for the specific role pairs as in Table 5. Note that we shall allow end users to provide the weights.

Table 5. An example of weights for specific roles

Role/co-role pair)	*Weight*
is a (Subtype of)/supertype of	0.8
is a (Subclass)/superclass of	0.8
Part-of/contain	0.5
Require/is required by	0.5
Has/belongs to	0.5
Has property/is property of	0.2
Equivalent to/equivalent to	1
Has member/is member of	0.5
Use/is used by	0.1

Note that the roles in a topic are not used for the matching. Only the roles are used to connect two terms from two topics are used for the matching. In other words, these roles and relevant lexons belong to the domain ontology Ω, but not ζ. For instance,

$$l_6 = \langle \gamma, senior\ employee, is\ a, supertype\ of, person \rangle \in \Omega$$
$$l_7 = \langle \gamma, junior\ employee, is\ a, supertype\ of, person \rangle \in \Omega$$
$$l_6, l_7 \notin \zeta_1 \cup \zeta_2 \cup \zeta_3 \cup \zeta_4 \cup \zeta_5$$

Then, the topics $\zeta_1, \zeta_2, \zeta_4, \zeta_5$are similar because l_1, l_2 contain the lexon term "person", which is linked to "senior employee" from l_4 by l_6 and linked to "junior employee" from l_5 by l_7. In particular, the similarity between l_1 and l_4 is 0.8 (as same as between l_2 and l_4, l_1 and l_5, l_2 and l_5) because "is a"/"supertype of" role pair has the weight of 0.8 in Table 5.

Including lexon terms and roles, the constraints in the commitment layer of Ω are the third component used for the semantic matching. For instance, the constraint *exclusive-or* draws a similarity of 0 between two lexons from two topics.

In the above discussion, the matching happens between two topics. It can as well happen between a topic and a set of topics. For example, we can search a set of relevant learning materials that can be used to enhance the assessment capacity "Coaching for performance". In this case, we need to find out for each lexon in this topic, whether it exists in other topics or not, using the same matching strategy as discussed above.

In this chapter, we have discussed principles of using topical ontology for the semantic matching. In the next chapter, we will illustrate the results.

5 Result

The methodology is implemented in a tool called PAD-ON Suite (Fig. 4). Users are able to create the patterns and specify the elements with concepts defined in the ontological basis and ontological assertions.

With PAD-ON suite, we have created in total 496 lexons, which are used and grouped in 49 ontological topics (Table 6). These topics include 10 assessment capacities, 9 OECD keys and 30 learning materials Table 7 shows the similarities between topics.

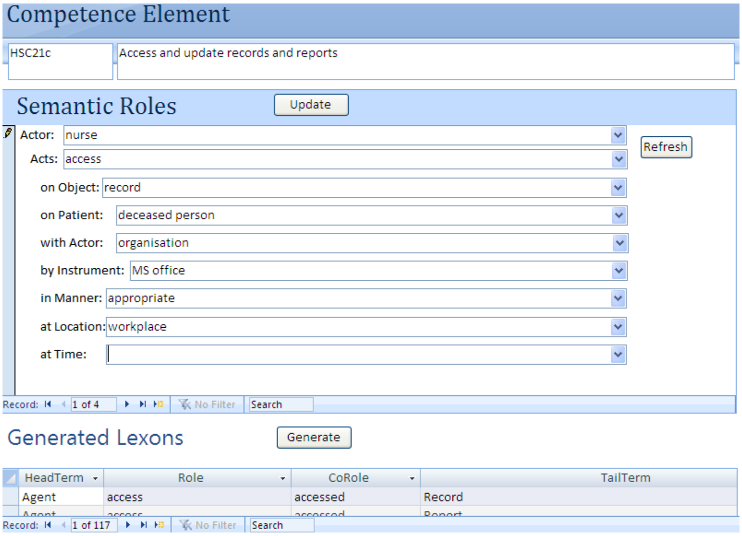

Fig. 4. PAD-ON suit (Screenshot)

Table 6. Number of lexons in 49 topics (data is collected on July 5, 2009)

Ontological topic name (selected)	*Number of lexons*
Assessment capacity	
Trustworthy	56
Helpful	43
Inspiring	19
Heart	12
Coaching for performance	39
Bottom line	27
Drive for results	84
Customer connected	54
…	…
Learning material	
Fault handling introduction system (FH 1)	11
Fault handling escalation (FH 2)	20
…	….

Table 7. Similarites between topics (selected)

Topic pair	*Similarity*	*Topic pair*	*Similarity*
Trustworthy, FH 1	0.38	Trustworthy, FH 2	0.21
Heart, FH 1	0.24	Heart, FH 1	0.22
…	…	…	…

6 Related Work and Discussion

Pattern-driven methodologies for creating domain ontologies are used to (semi-) automatically create ontologies by populating concepts in patterns. A related work of this paper is OntoCase [2], which retrieve and select patterns from a set of ontological patterns. The patterns that get high ranks are further used by ontology-based applications, such as search engines.

Chenine et al. [3] use context based ontology design architecture in the form of Principle-Subject-Support (PSS) pattern, for creating distributed, heterogeneous and multi-functioned, application ontology. It is applied to create a military domain ontology.

Presutti and Gangemi [10] discuss how to extract and describe emerging content ontology design patterns (CP), and how to compose, specialize and expand them for ontology design. They illustrate two CPs and apply them in the music industry domain. Other related work can be found in [1, 4, 11].

7 Conclusion and Future Work

In this paper, we have discussed the design of a pattern-driven topical ontology modeling methodology (PAD-ON), which is supported by a tool called PAD-ON suit. The resultant ontology is used by the semantic matching engine.

In the future, we will design and implement a more generic matching framework, which contains many matching strategies from different perspectives, such as linguistic, relational database modeling and graph theory.

Acknowledgement. This paper is support by the EC Prolix project.

References

1. Aranguren, M.E., Antezana, E., Kuiper, M., Stevens, R.: Ontology Design Patterns for bio-ontologies: a case study on the Cell Cycle Ontology. BMC Bioinformatics 9(5), S1 (2008)
2. Blomqvist, E.: Pattern ranking for semi-automatic ontology construction. In: Proceedings of the 2008 ACM symposium on Applied computing table of contents, Fortaleza, Ceara, Brazil, pp. 2248–2255 (2008) ISBN:978-1-59593-753-7
3. Chenine, M., Kabilan, V., Lozano, M.G.: A Pattern for Designing Distributed Heterogeneous Ontologies for Facilitating Application Interoperability. In: Proceedings of 3rd Open INTEROP Workshop On Enterprise Modeling and Ontologies for Interoperability (EMOI 2006), collocated with CAiSE 2006 (2006)
4. Clark, P., Thompson, J., Porter, B.: Knowledge patterns. In: Cohn, A.G., Giunchiglia, F., Selman, B. (eds.) Proceedings of KR 2000: Principles of Knowledge Representation and Reasoning, pp. 591–600. Morgan Kaufmann, San Francisco (2000)
5. Fellbaum, C.: WordNet: an electronic lexical database (Language, Speech, and Communication). MIT Press, Cambridge (1998)
6. Gruber, T.R.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: Guarino, N., Poli, R. (eds.) Workshop on Formal Ontology, Padva, Italy. In Book, Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Dordrecht (1993)
7. Levenshtein, VI: Binary codes capable of correcting spurious insertions and deletions of ones (original in Russian). Russian Problemy Peredachi Informatsii, 12–25 (1965)

8. Meersman, R.: Ontologies and Databases: More than a Fleeting Resemblance. In: d'Atri, A., Missikoff, M. (eds.) Proceedings of OES/SEO 2001 Rome Workshop. Luiss Publications (2001)
9. Presutti, V., Gangemi, A.: Content Ontology Design Patterns as Practical Building Blocks for Web Ontologies. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 128–141. Springer, Heidelberg (2008)
10. Rector, A., Rogers, J.: Patterns, properties and minimizing commitment: Reconstruction of the GALEN upper ontology in owl. In: Gangemi, A., Borgo, S. (eds.) Proceedings of the EKAW 2004 Workshop on Core Ontologies in Ontology Engineering (2004)
11. Sowa, J.F.: Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley, Reading (1984)
12. Spyns, P., Meersman, R., Jarrar, M.: Data Modeling versus Ontology Engineering. SIGMOD Record: Special Issue on Semantic Web and Data Management 31(4), 12–17 (2002)
13. Zhao, G., Meersman, R.: Architecting Ontology for Scalability and Versatility. In: Meersman, R., Tari, Z. (eds.) OTM 2005. LNCS, vol. 3761, pp. 1605–1614. Springer, Heidelberg (2005)