RP-Match: A framework for Automatic Mapping of Regulations with Organizational Processes

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Krishna Sapkota[a,*], Arantza Aldea[a] , Mohammad Younas[a,] David A Duce[a] and Rene Banares-Alcantara[b]
[a] *Department of Computing, Oxford Brookes University, Oxford, UK*
[b] *Department of Engineering Science, University of Oxford, Oxford, UK*

Abstract. Mapping organizational processes with applicable regulatory guidelines is an important step in Regulatory Compliance Management. Automation in the mapping process helps in automation of the overall compliance process. Although there are several approaches which compute mapping between different entities such as ontology mapping, sentence similarity, semantic similarity and regulation-requirement mapping; it still requires a framework to automate the mapping process between regulation and processes. In this paper, we present RP-Match framework, which exploits the state of art tools for similarity measures and proposes cascading-priority algorithm for the regulation-process similarity computation. The framework also avails from the structures of the regulation ontology and organizational process ontology for the similarity measure. An initial case study carried out in the Pharmaceutical industry has shown an encouraging result.

Keywords: Regulation-Process Mapping, Regulatory Compliance Management, Semantic Similarity, WordNet Similarity

1. Introduction

Regulatory Compliance Management (RCM) is a management process, which an organization implements to ensure organizational compliance with relevant requirements and expectations. Examples of requirements are regulations or legal guidelines, and that of expectations are mandates, policies and guidelines for the organizations. It poses a strain to organizations as on the one hand, non compliance to regulations results in heavy penalties, companies being sued or even suspension and closure and on the other hand, maintaining compliance with the regulations also costs heavily. Such an example is presented in [1], which concluded that legislations such as the Sarbanes-Oxley Act (SOX) imposed stringent compliance requirements and organizations had to make heavy investments to meet the requirements.

Checking and managing the compliance manually is an arduous, extensive and error prone task. It necessitates expertise in the field and their enormous amount of time, which costs a huge amount of capital investment to organizations and is still error prone. In order to automate the process and make it errorless, computer aided CM systems have been developed [1–3]. However, these systems are still experiencing various challenges to streamline and automate the process. The key challenge being experienced by the current systems is that they have difficulty on keeping up with the frequently changed and updated regulations as updating the system manually consumes great amount of time, efforts and investment. In the current systems, the requirements and processes are embedded within the specific codes designed for a specific domain for a specific purpose. Systems developed by mixing the codes with the regulatory

*Corresponding author. E-mail: ksapkota@brookes.ac.uk

knowledge and organizational processes are hard to trace, update and reuse. Representing the regulatory knowledge and organizational process knowledge in standard, homogeneous and interoperable formats can improve updating processes and reusability. It can be reified by representing the knowledge in semantic web technologies.

In particular, modeling the organizational processes and regulations governing these processes in process-ontology and regulation-ontology helps in the reusability of the knowledge. In [4], besides representing regulations and processes in semantic formalism, a special type of concept called *ValidationTask* has been created. It is a concept connecting regulations with the processes, and is associated with the ontological concepts and properties where application of the inference engine can deduce whether or not a process complies with a regulation. However, the representations of the processes, regulations and design of the validation tasks need to be changed or updated in circumstances such as 1) change/update in the existing regulations or 2) the processes need to conform to regulations from other regulatory bodies or in other territories. In such cases, mapping of the new regulations with the processes and validation tasks constitutes an important step towards updating the affected processes and validation tasks. Furthermore, automation in the mapping process also contributes in the overall automation of RCM.

Among the challenges in the process of automatic mapping between regulatory and process entities are 1) ambiguity and complexity of the regulatory text, 2) implicit information in the description of organizational processes and 3) absence of a standard framework to work with regulation and process ontologies in order to facilitate the processes such as mapping. Among these challenges, this paper focuses on the third challenge. In particular, it proposes a framework to map regulations with processes and validation tasks.

In order to address the mapping challenge, the techniques used to compute semantic relatedness among entities such as sentence, word and concept have been useful. In particular, sentence to sentence mapping [5–8], concept to concept mapping [9–11] and word to word mapping [12] can be regarded as solutions to the problem to some extent. In sentence to sentence mapping, two sentences are compared to determine how much they are related. Concept to concept mapping or ontology mapping uses several properties of the concepts in order to compute the similarity. Likewise, in word to word mapping or word similarity uses lexical onology such as WordNet in order to find the similarity and distance in the meaning of the words.

However, in process domain where the processes are modeled in to an ontology and validations tasks have been defined for the compliance checking, there requires a mapping process which considers the structure of the validation tasks and the semantics embedded in the regulatory text [13].

As a solution to the automation in the mapping of regulations with processes and validation tasks, this paper presents RP-Match framework. This is an extension of the mapping algorithm presented in [13], which assumes that the regulations and processes are represented in ontologies. The proposed framework has been developed as a part of a larger research work towards developing a semantic compliance management system, which aims to contribute on automating 1) extraction and representation of regulatory-entities and 2) mapping regulations with processes. This paper concentrates on the second issue with RP-Match framework as a solution.

In the regulation side, this framework considers the core entities such as subject and action, entity modifiers or auxiliary entities such as place, time, purpose and conditions, and topics of the regulations. In the process domain, it considers the subject, action, process and validation tasks. The summary of the contribution of this paper are 1) automating the mapping process between regulations and processes and 2) use of cascading priority algorithm for similarity, where the priority of the similarity decreases from topic similarity to core-entity similarity and to auxiliary-entity similarity.

The rest of this paper is organized as follows. A literature review of related work is shown in section 2, the RP-Match framework is described in Section 3. Section 4 analyses the results obtained from the case study and finally Section 5 concludes the paper with a summary and future work.

2. Related Work

There are several approaches in requirement engineering, domain about requirement-process alignment and requirement-regulation mapping. In requirement-process alignment, stakeholder's requirements are elicited, and the processes needed to fulfill the requirements are indentified. In this process, metadata for the requirement and processes are created, and similarity between them is computed. In requirement-regulation mapping, similarity between

requirements and regulatory guidelines are computed using various approaches such as sentence mapping and metadata similarity. However, the legal domain is still in need of efficient algorithms to map regulations with processes.

All the state-of-art similarity algorithms can be considered as related to this work such as sentence similarity, word similarity, ontology mappings and conceptual distance. In [5], sentence similarity is computed using align-heuristics where noun, verb, adjective, adverb and numbers are aligned, and the approach was inspired by the popular sentence alignment algorithm in [7]. In this framework, the decomposition of sentences into different entities for the similarity measure is similar to the proposed framework; however this can be only applied to compute sentence similarity. RP-Match can be applied to sentence to the ontological concept matching considering selected properties of the concepts. Similarly, sentence matching based on BoW algorithm with emphasis on the important terms are applied in [6] in order to determine the answer similarity. The way BoW similarity has been implemented is related to this work; however, it is only applicable to compare sentences. A pilot for similarity in SemEval competition has described the similar algorithms for the sentence similarity which also requires training and testing sentences[8].

Conceptual distance and similarity computation in ontologies are also related to this work. The use of weight allocation and node routing table in order to compute semantic distance between two concepts in an ontology [11] is related to the proposed framework. However, the measure can only be applied to two concepts in an ontology; whereas RP-Match can be implemented to determine the similarity between a concept in an ontology with a sentence represented in various concepts in another ontology. Similarly, in [10], graph based similarity is computed considering various types of ontological properties and depth of the concepts. The similarity based on the depth and max depth has been considered for the future work of the current framework in order to extend the current difference table. In [9], two ontologies have been defined in order to determine similarity of a new event with the existing event. The similarity computed using WordNet similarity is related to this work; however, it requires that both ontological concepts and individuals manually designed by the domain expert. In RP-Match, only the individuals in process ontology are manually designed by the system engineers.

Application of a combination of similarity approaches in [14] in order to determine similarity between contents of two television programs have been found similar to the proposed framework. The most related part in this framework is the computation of the similarity of topics and the text in the synopsis. In particular, BoW similarity and hierarchical similarity applied for the purpose are related to this work. However, it is only applicable if both compared entities contain hierarchy and text description in sentences. In the mapping between regulation and processes, the regulations are represented in fractioned sentences and processes are represented in ontological concepts, therefore, it cannot be applied in this mapping. The RP-Match can be applied to determine similarity where one side is represented as ontological concepts with strictly defined properties, and the other side with some algorithm to extract meaningful entities from the text [15], or text is already represented in a meaningful representation format such as ontology.

3. The Framework

The RP-Match framework aims to contribute as a step towards the automation of the overall CM, by reducing the manual work in mapping the regulations with validation tasks or processes. Mapping is a process which identifies processes and validation tasks affected by a change in a regulation. It is a semi-automatic process, which generates a mapping table, and a domain expert has to verify the recommendations [13]. It assumes that there exists a process ontology which represents organizational processes and validation tasks. It also requires that the regulations are represented in an ontology [15] so that the regulatory-entities such as topics, core-entities and auxiliary entities can be used for the mapping process.

There are three similarities computed in this framework, and they are 1) topic similarity 2) core-entity similarity and 3) auxiliary-entity similarities. It is an extension of the previous framework [13] with considerations of two more similarities i.e. topic-similarity and auxiliary-entity similarity. The cascading heuristics of overall similarity measurement algorithm prioritizes the similarity measures from top to bottom as topic, core-entity to auxiliary entity. In each similarity measure, the ontological structures of the regulation and process ontologies are exploited,

where the important ontology axioms such as "disjointWith" is considered for the difference and similarity score computation of two concepts. Figure 1shows how three types of similarity are computed and Figure 2 depicts a mapping between a regulation and a validation task in SemReg and OntoReg ontology.

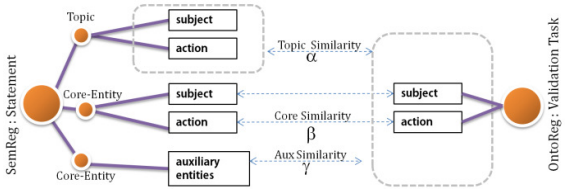


Figure 1. Three different types of similarity computations in RP-Match framework

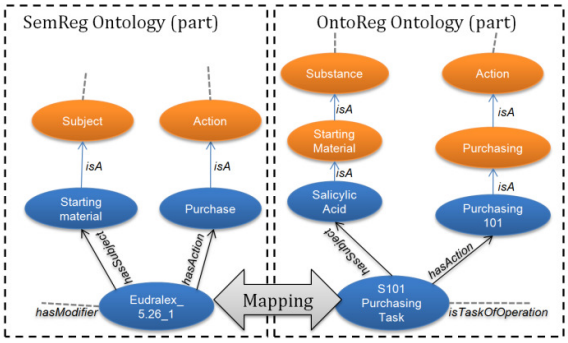


Figure 2. Mapping regulation with validation task using regulation and process ontology

3.1. Conceptual distance computation

In the similarity computation, the proposed framework determines whether a concept or an individual in the regulatory ontology is similar to a concept or individual in process ontology. Although some concepts look like very similar to each other in a general context, a domain-ontology may have defined the concept as different from each other. In this case, a lexical ontology such as WordNet deduces that these concepts are similar; however at the same time domain ontology such as process ontology declares that they are different from each other. For an example, as shown in Figure 3 the concepts "substance" and "equipment" are more related in WordNet; where as in process ontology, they are defined as different from each other. Therefore, it is necessary to determine how different two concepts are from each other within a domain-ontology in order to define similarity.

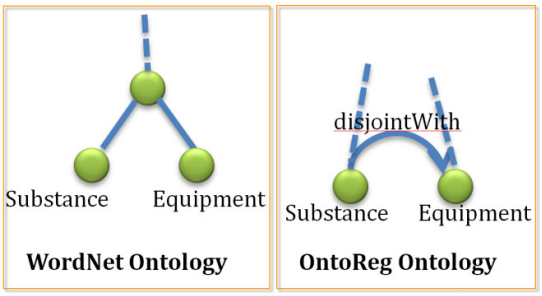


Figure 3. Different ontologies showing similarity and differences between the same concepts

The distance between the two concepts in the process ontology is computed considering the axiom "disjointWith". Currently, the value becomes 1 or 0 considering if they are disjoint with each other or not respectively; but in the future, we aim to consider the semantic distance computation algorithm [11] in conjunction with the current algorithm. After the conceptual difference computation, a table is created, in which each row is represented by <C1, C2, d>, where C1 and C2 are two concepts and d is the difference value.

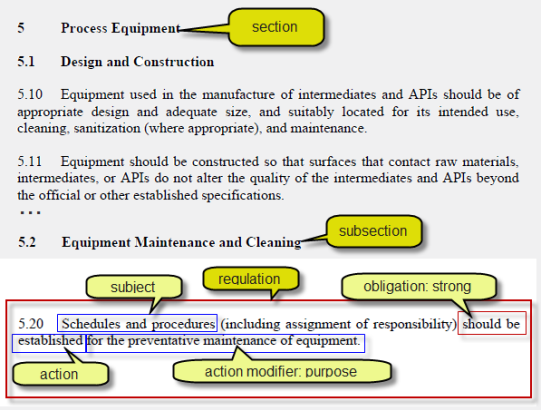


Figure 4. An excerpt from Eudralex regulation showing regulatory entities

3.2. Similarity score computation

In a regulation ontology, regulations (Reg) are organized under topic hierarchy (Topic) such as part, chapter, section and subsection. By term "regulation", we referred to a unit of regulatory guidelines, which is usually represented in a paragraph. Each regulation may be made up of one or two regulation statements (Stmt). A statement is a sentence in a regulation.

Each statement comprises core-entities (Core) and auxiliary-entities (Aux). The core-entities represent subject (Sub) and action (Act), and the auxiliary-entities represent extra information such as time, place and purpose. An example of the regulatory text depicting topics such as section and subsection, core-entities such as subject and action and auxiliary entities such as action modifier is presented in Figure 4.

In a process ontology, a validation task (Task) is a special concept which helps in compliance checking of an organizational process. The two most important Concepts in the validation tasks are associated with are subject (Sub) and action (Act). A process contains multiple unit procedures and operations [4]. An operation is a unit of a process; however it depends on how the processes are modeled by the domain experts. Each operation is associated with various essential entities such as equipments, substances, documents, responsible person, time, written procedure, reports and validation task.

Table 1. Pseudocode to compute subject similarity

Input: *stmt* is a regulation and *task* is a validation task.
Output: *subject_score* is the similarity score between the subjects.

function GET-SUBJECT-SCORE(*stmt, task*) **returns** *subject_score*

1: *subject_score = 0*
2: $S_1 = \{s_1 \mid s_1$ is_a_subject_in *stmt*$\}$
$S_2 = \{s_2 \mid s_2$ is_a_subject_in *task*$\}$
3: **for each** $s_i \in S_1$
4: **for each** $s_j \in S_2$
5: d ← GET-DIFFERENCE-VALUE(s_i, s_j)
6: **if** ($d < \theta$) **then**
7: *new_subject_score* ← COMPUTE-WORDNET-SIMILARITY-SCORE(s_i, s_j)
8: **if** (*new_subject_score* > *subject_score*) **then**
9: *subject_score* ← *new_subject_score*
10: **end if**
11: **end if**
12: **end for**
13: **end for**
14: **return** *subject_score*

Description: This function computes a similarity score between the subjects of a regulation-statement and a validation-task in regulation- ontology and process-ontology respectively. In this process initially the score is set as 0, which will be updated with the computed value. Let's take two sets of subjects, i.e. S_1 from the regulation-statement and S_2 from the validation-task. Now, we compare each word in these sets in order to compute the difference-value d from the difference-table created from the process-ontology. If the two words are not defined as ontologically different in the process-ontology, d will be below the threshold θ, and a state-of-art similarity algorithm based on WordNet lexical ontology (Lin similarity, in this case) is applied. It results into a set of similarity-scores, from which the highest similarity-score is selected as the final similarity-score of the subjects between the statement and the task.

In the proposed cascading-priority algorithm, three types of similarities are computed, and they are 1) Topic similarity (Topic Vs. Task), 2) Core-entity similarity (Core Vs. Task) and 3) Auxiliary-entity similarity (Aux Vs. Task). Figure 1 displays how the three similarities are computed in the cascading-priority algorithm.

In the topic similarity, the similarity score is computed using the Bag of Word (BoW) similarity algorithm [14]. In this process, each word in the topic text is compared with subject and action of each validation task or each operation. The similarity of a topic with a validation task or an operation is determined by the

percentage of the words in a topic that is similar to the validation task or the operation.

In the core-entity similarity, average of the subject and action similarity is computed. In particular, the similarity scores are computed for subject and action separately. Each individual in a statement representing as a subject in the regulatory ontology is compared with that in a validation task or an operation are compared for the similarity measure. In the process ontology, for the subject similarity computation, the individuals and concepts associated with validation tasks or operations are analyzed. The subjects in the validations tasks or operations are equipment, substance, operation, document and personnel. Table 1 demonstrates the pseudocode to compute the similarity score between the subjects.

Similarly, the action similarity is computed by comparing the action words associated with a regulatory statement and a validation task or an operation. When these two similarity scores are computed, an average of these two score is determined as the similarity score of the core-entity matching i.e. Sim = (subject similarity + action similarity)/2.

In the auxiliary-entity similarly, the similarity score is computed considering the similarity between each word in the auxiliary-entities and the subject and action of a validation task or an operation by using BoW [14] algorithm similar to the topic similarity computation.

In all the similarity computation processes above, the two compared terms are checked against the "Difference Table" computed in the previous step in order to determine whether they are defined as different in the process ontology. It also requires setting a threshold θ; if a difference value exceeds the threshold, the compared terms are regarded as different

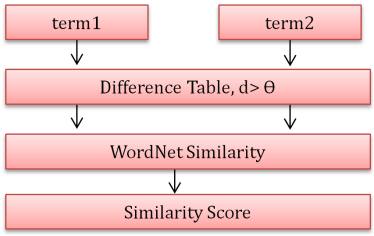


Figure 5. The similarity calculation process with consideration of difference table

from each other, and they will not be further processed for the similarity score. Figure 5 depicts a similarity computation process between two words where θ is set as 0.5. When these terms are below the threshold, a state of semantic similarity measure, Lin similarity [16] is applied. It is based on the nodes in a lexical ontology such as WordNet and information content values of the concepts. The advantage of using this measure is that it considers the hierarchical structure of the dictionary meaning in the lexicon and generality of the concepts in large generic corpora. The computed similarity values are generated as a real number between 0 and 1.

3.3. Cascading priority algorithm

After the similarity scores computation of three types, it needs to determine the overall similarity of the regulation-statement with a validation task or an operation. This can be achieved by computing a weighted sum of the three similarity values. The challenge remains in assigning a weight to a similarity value. In order to tackle this challenge, we propose an algorithm called cascading priority algorithm. In this algorithm, the priorities of the similarities are given to the top levels as compared to the lower levels. In particular, there is more priority on the topic-similarity compared to core-entity similarity, and again the core-entity similarity has more priority than auxiliary-entity priority. The algorithm is depicted and explained in more detail in Table 2

Table 2. Cascading-Priority algorithm

Let α, β and γ be the weights for the topic, core and auxiliary similarity scores, and Sim_{Topic}, Sim_{Core}, Sim_{Aux}, Sim_{All} and be the topic similarity, core-entity similarity, auxiliary-entity similarity and overall similarity respectively. From the repeated observations, the optimum values of α, β and γ are found as 0.9, 0.8 and 0.7 respectively. The following heuristics has been applied in order to compute the weighted sum of the similarity scores.

(1) If $Sim_{Core} < Sim_{Topic}$ & $Sim_{Aux} < Sim_{Topic}$

$Sim_{All} = \alpha \times Sim_{Topic} + (1-\alpha) \times MAX(Sim_{Core}, Sim_{Aux})$

If the topic similarity has the highest score, the highest weight α is assigned to the topic-score. From the rest, the highest score is selected, and a complementary weight with respect to α is assigned in order to make the total weight 1.

(2) If $Sim_{Topic} < Sim_{Core}$ & $Sim_{Aux} < Sim_{Core}$

$Sim_{All} = \beta \times Sim_{Core} + (1-\beta) \times MAX(Sim_{Topic}, Sim_{Aux})$

If the core similarity has the highest score, the second highest weight β is assigned to the core-score. The maximum score between the topic and auxiliary

similarity is selected and the complementary value of β is assigned to the score.

(3) If $Sim_{Topic} < Sim_{Aux}$ & $Sim_{Core} < Sim_{Aux}$

$Sim_{All} = \gamma x\, Sim_{Aux} + (1- \gamma) x\, MAX\,(Sim_{Topic}, Sim_{core})$

If auxiliary similarity has the highest score, the lowest weight γ is assigned to the score. The maximum value between the topic and core scores is selected and the complementary value of γ is assigned to the score.

3.4. Statement similarity to regulation similarity computation

As mentioned earlier, each regulation is composed of one or more statements. The overall similarity computed in the step above is the similarity of a statement with a validation task or an operation in the process ontology. Now, if a regulation contains more than one statement, it also contains a set of similarity scores and the maximum score in the set is regarded as the similarity score of the regulation i.e. $Sim_{Reg} = MAX(Sim_{s1}, Sim_{s2},...,Sim_{sn})$.

4. Result and Evaluation

For the case study, the Pharmaceutical industry in EU was chosen after considering it as one of the heavily regulated domains. The regulation governing this domain in EU is the Eudralex [17]. Likewise, for the process-ontology in the case study, OntoReg ontology was borrowed from Engineering Science Department in the University of Oxford [4]. A regulatory-ontology, SemReg was created by extending LKIF-Core ontology [15]. The concepts of SemReg were populated with the extracted regulatory-entities. An excerpt of SemReg ontology is depicted in Figure 6.

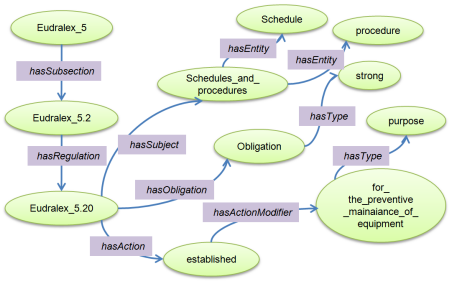


Figure 6. An excerpt of SemReg ontology showing regulation representation

Table 3. Result of RP-Match framework showing mapping score between regulation and validation task

SN	Regulation	Validation Task	Mapping Score	Accepted/ Rejected	Evaluation
1	Eudralex_V1_5.21	FilterCleaningTask	1.0	Accepted	TP
2	Eudralex_V1_5.21	FilterCleanlinessTask	0.94	Accepted	TP
3	Eudralex_V1_5.21	T101CleaningTask	1.0	Accepted	TP
4	Eudralex_V1_5.21	T101CleanlinessTestTask	0.94	Accepted	TP
5	Eudralex_V1_5.21	T102CleaningTask	1.0	Accepted	TP
6	Eudralex_V1_5.21	T102CleanlinessTestTask	0.94	Accepted	TP
7	Eudralex_V1_5.22	FilterCleaningTask	1.0	Accepted	TP
8	Eudralex_V1_5.22	FilterCleanlinessTask	0.94	Accepted	TP
9	Eudralex_V1_5.22	T101CleaningTask	1.0	Accepted	TP
10	Eudralex_V1_5.22	T101CleanlinessTestTask	0.94	Accepted	TP

The OntoReg ontology was provided with five regulations from the Eudralex 2007 which are associated with twelve validation tasks designed by domain experts. We extended it with more regulations and validation tasks in order to measure true positive (correct mapping), false positive (wrong mapping) and false negatives (missing mapping). We decided to obtain initial results based on these regulations and the validation tasks, as a bench mark for accuracy of the proposed framework. The regulations in SemReg are also populated with Eudralex 2007 (Eudralex-V1) and Eudralex 2009 (Eudralex_V2) and the proposed framework is applied to map the regulations in SemReg ontology with the validation tasks in OntoReg ontology. Table 3 and Table 4 show the mapping result obtained after application of the RP-Math framework. The mapping result is compared with the result of the previous mapping algorithm, and precision, recall and f-measure of the result are computed based on the true positive (TP), false positive (FP) and false negative mappings (FN).

Table 4. Result showing correct, incorrect and missing mappings

Type of mapping	No of mappings
Correct Mapping (TP)	164
Incorrect Mapping (FP)	19
Missing Mapping (FN)	36

The precision and recall are computed using the formula, $Precision = \frac{T}{TP}$, $Recall = \frac{T}{TP}$. Figure 7 shows the accuracy of the proposed framework as compared to the previous framework.

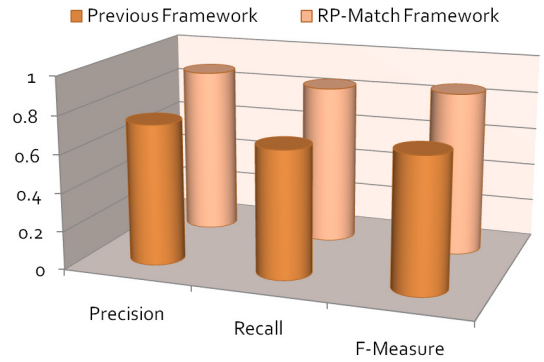


Figure 7. Accuracy comparison of RP-Match framework with previous framework

5. Conclusion and Future Work

Mapping regulations with organizational processes is important particularly when there are changes in the regulations, or an organizational process needs to follow regulations from different policy makers. Although there have been state-of-art algorithms for the mapping between requirements and processes in requirement engineering domain; there still needs to have efficient algorithms to map regulations with processes. Several similarity measuring algorithms have potential components, which can be used to streamline the mapping process. Having considered the above facts, we presented RP-Match framework for the mapping process. The process computed three typed of similarity scores; 1) topic similarity, 2) core-entity similarity and 3) auxiliary-entity similarity. An innovative algorithm called Cascading-Priority has been applied in order to calculate the weighted sum of these three similarity scores. The framework also considered the ontological axiom "distinct From" to create a difference look up table for the similarity score computation. The initial result from a case study in Pharmaceutical industry has demonstrated an encouraging result.

At present, the processed ontology was equipped with very few regulations and validations tasks, based on which the current result was compared and analyzed. In the future, we aim to compare the results with more regulations and validation tasks in the currently used regulations i.e. Eudralex 2007. Furthermore, we also aim to test the framework with regulations and processes in different domains

References

[1] S. Ghanavati and D. Amyot, "Towards a framework for tracking legal compliance in healthcare," *Proceedings of the 19th international conference on Advanced information systems engineering CAiSE'07*, pp. 218–232, 2007.

[2] S. I. Haider, *Validation standard operating procedures*. New York: CRC Press, 2006, p. 1049.

[3] Y. Liu, S. Muller, and K. Xu, "A static compliance-checking framework for business process models," *IBM Systems Journal*, vol. 46, no. 2, pp. 335-361, 2007.

[4] M. B. Sesen, P. Suresh, R. Banares-Alcantara, and V. Venkatasubramanian, "An ontological framework for automated regulatory compliance in pharmaceutical manufacturing," *Computers & Chemical Engineering*, vol. 34, no. 7, pp. 1155-1169, Jul. 2010.

[5] D. McCarthy and S. Gella, "DSS: Text Similarity Using Lexical Alignments of Form, Distributional Semantics and Grammatical Relations," in *First Joint Conference on Lexical and Computational Semantics*, 2012, pp. 557-564.

[6] M. Mohler and R. Bunescu, "Learning to grade short answer questions using semantic similarity measures and dependency graph alignments," *of the 49th Annual Meeting of*, 2011.

[7] R. Barzilay, "Sentence alignment for monolingual comparable corpora," *Proceedings of the 2003 conference on*, 2003.

[8] E. Agirre, D. Cer, and M. Diab, "Semeval-2012 task 6: A pilot on semantic textual similarity," *(SemEval 2012), in*, no. 3, pp. 385-393, 2012.

[9] Y. Chen, J. Wang, and Z. Cheng, "An algorithm to compute similarity between danger objects based on ontology for danger-aware systems," *Aware Computing (ISAC*, pp. 128-135, 2010.

[10] A. Hawalah, "A graph-based approach to measuring semantic relatedness in ontologies," *Proceedings of the International Conference on*, 2011.

[11] J. Ge and Y. Qiu, "Concept Similarity Matching Based on Semantic Distance," *2008 Fourth International Conference on Semantics, Knowledge and Grid*, no. C, pp. 380-383, Dec. 2008.

[12] T. Pedersen and J. Michelizzi, "WordNet□:: Similarity - Measuring the Relatedness of Concepts," *Architecture*, 2004.

[13] K. Sapkota, A. Aldea, D. A. Duce, M. Younas, and R. Banares-Alcantara, "Towards Semantic Methodologies for Automatic Regulatory Compliance Support," in *Proceedings of the 4th workshop on Workshop for Ph.D. students in information & knowledge management (PIKM '11)*, 2011, pp. 83-86.

[14] Z. Yu and X. Zhou, "Combining Vector Space Model and Category Hierarchy Model for TV Content Similarity Measure," *2009 Third International Conference on Multimedia and Ubiquitous Engineering*, pp. 130-136, Jun. 2009.

[15] K. Sapkota, A. Aldea, M. Younas, D. A. Duce, and R. Banares-Alcantara, "Semantic-ART: a framework for semantic annotation of regulatory text," in *Proceedings of the fourth workshop on Exploiting semantic annotations in information retrieval (ESAIR '11)*, 2011, pp. 23-24.

[16] D. Lin, "An information-theoretic definition of similarity," *Proceedings of the 15th international conference*, 1998.

[17] Eudralex, "The Rules Governing Medicinal Products in the European Union," *Online*, 2011. [Online]. Available: http://ec.europa.eu/health/documents/eudralex/cd/index_en.htm. [Accessed: 18-Jul-2011].