

A Semantic-enabled Framework for Drug Recommendations

Charalampos Doulaverakis¹, George Nikolaidis², Athanasios Kleontas MD^{2,3},
and Ioannis Kompatsiaris¹

¹ Information Technologies Institute, Centre for Research and Technology Hellas,
Thessaloniki, Greece, email: {[doulaver](mailto:doulaver@iti.gr),[ikom](mailto:ikom@iti.gr)}@iti.gr

² Ergobyte S.A., Thessaloniki, Greece, email: {[gnikolaidis](mailto:gnikolaidis@ergobyte.gr),
[kleontas](mailto:kleontas@ergobyte.gr)}@ergobyte.gr

³ AHEPA, University General Hospital of Thessaloniki, Greece

Abstract. The paper presents Panacea, a semantic-enabled framework capable of offering drug-drug and drug-diseases interaction discovery. For enabling this kind of service, medical information and terminology had to be translated to ontological terms and be appropriately coupled with medical knowledge of the field. International standards for diseases and drug classifications provide the backbone of the common representation of medical data while the medical knowledge of drug interactions is represented by a rule base which makes use of the aforementioned standards. The framework provides an efficient drug recommendations service where Semantic Web technologies are coupled with traditional business rule engines.

Keywords: ontologies, medical information modelling, rule-based reasoning

1 Introduction

One of the health sectors where intelligent information management and information sharing compose valuable preconditions for the delivery of top quality services is personalized drug prescription. This is more evident in cases where more than one drug is required to be prescribed, a situation which is not uncommon, as drug interactions may appear. The problem is magnified by the wide range of available drug substances in combination with the various excipients in which the former are present. Semantic Web technologies can play an important role in the structural organization of the available medical information in a manner which will enable efficient discovery and access. Research projects funded for enabling Semantic Web technologies in the diagnosis and therapeutic procedures exist such as REMINE (<http://www.remine-project.eu/>) and PSIP (<http://www.psip-project.eu/>) or works such as [1], but they don't fully address the problem of automated drug prescription using drug-drug and drug-disease interactions.

In this work Panacea, a semantic-enabled system for discovering drug recommendations and interactions, is presented. Panacea is based on experiences and

lessons drawn from the development of GalenOWL [2], a similar system which had Semantic Web technologies in its core. As such, Panacea can be considered the evolution of GalenOWL in terms of design and scalability. The paper is organized with Section 2 giving details of the framework’s architecture, the semantic modelling and reasoning approaches, while Section 3 evaluates the framework, discusses the results and concludes the paper.

2 Architecture, semantic modelling and reasoning

In order to demonstrate the benefits of the proposed drug recommendation system, a use case is described: An elder man visits his family doctor complaining for pain in his abdominal region which is accompanied with fever and he is diagnosed with right pyelonephritis. The patient is already receiving clopidogrel for chronic atrial fibrillation, and candesartan and amlodipine for high arterial blood pressure. For treating the patient, the doctor has to prescribe drugs in order to treat the resulting inflammation and fever, the cause of the inflammation (the pyelonephritis) and the abdominal pain. However, there should be checks for interactions of the drugs that the patient is already prescribed with the new condition and also verify that the new prescription will not have adverse effects or interactions with the patient’s current medication and medical history. It is clear that the task for the doctor is hard and using an automated drug recommendation system can minimize this risk of wrong prescriptions.

The purpose of Panacea is to provide drug prescription recommendations based on a patient’s medical record and drug prescription rules. For details regarding the initial medical and pharmaceutical data that were available, the reader is encouraged to read [2].

Panacea follows a layered reasoning process. During initialization, standard medical terminologies and custom encodings are transformed to semantic entities, using an appropriate vocabulary, and the initial ontology is constructed. A reasoner binds to the ontology to infer relations such as inheritance, unions and equivalences. In order to get recommendations in Panacea, a patient instance with the appropriate medical record data is created and is fed to the knowledge base. The reasoning process enriches the patient instance with inferred knowledge, thus making that knowledge explicit. On this enriched instance, and by utilizing a different reasoning process, the set of medical rules is applied upon. The result of this final stage of rule-based reasoning is the recommendations list which can be retrieved through querying.

Semantic transformations Panacea is built on top of international standards of medical terminology in order to represent medical and pharmaceutical information. The standard terminologies that are used are ICD-10 (International Classification of Diseases), ATC (Anatomical Therapeutic Chemical classification), UNII (Unique Ingredient Identifier) and IVT (International Virus Taxonomy). Besides these international standards, a number of domain classifications have been declared in order to enhance the usability of the system or to represent data that are not included in the standards. These are a) a *Substance*

classification which represent human-readable names for drug active substances and correspond to ATC codes, b) *Custom Concepts* where they represent diseases or drugs that are not present in the international standards, and c) *Custom Collections* where they represent “groups” of substances and diseases that are frequently present in drug interactions.

SKOS vocabulary: The medical classifications were expressed using the SKOS⁴ vocabulary. Each of the classifications is placed under the *skos:ConceptScheme* class and each entry is placed under the *skos:Concept* class. Classifications hierarchies are expressed with the *skos:broaderTransitive* and *skos:narrowerTransitive* properties while *Custom Collections* are represented with the *skos:member* property. Equivalences, e.g. between ATC codes and Substances, are expressed with the *skos:closeMatch* property. A total of 64, 658 definitions of classification codes have been expressed using SKOS.

Medical reasoning When querying the system for recommendations, a patient instance is created with the initial patient data and is loaded in the knowledge base. The reasoner, using RDFS inference and a small number of additional rules, infers all the implicit patient data. As an example, a patient who suffers from a form of thrombocytopenia (*icd:D69.6*) will be inferred as also suffering from *icd:D69*, *icd:D65-D69*, *icd:D50-D89* (through *skos:broaderTransitive*) and from *pnc-cc:deficiency-bone-marrow* (through *skos:member* from the *Custom Collection* definition).

Rule-based reasoning Drug recommendations in Panacea are generated using a rule-based approach. The rules express the indications and contraindications of drug substances while their premises are the medical definitions and the patients’ age and sex group. The rules use the logical operators *and* (&) and *or* (|) and parentheses. An example of a rule is for the substance “lisuride” which is expressed as “*lisuride* = *icd:E22.0* | (*icd:E22.1* & (*icd:N91.0* | *icd:N97*)), *ageGroup=adult* or *elder*”. The rule reads that: substance “lisuride” is recommended for adult and elder patients who suffer from E22.0, OR suffer from E22.1 AND one of the N91.0 OR N97.

The rule base consists of 1,342 rules whose validity has already been assessed in [2]. These rules have to be properly parsed and transformed in order to match the knowledge base and the, enriched with implicit knowledge, patient instance. Any rule reasoner can be utilized to use the rule base. To demonstrate this concept, two separate rule engine integrations, the Jena⁵ and Drools⁶ rule engines, have been developed. While the Drools integration only required to translate the rule base to the Drools Rule Language and develop an additional layer for Drools to interface with the Patient instance, the Jena integration proved more complex as the rule base is quite large to be handled efficiently and certain optimizations had to be implemented. These optimizations involve the introduction of coarse rule selection phase where a subset \mathcal{A} of candidate rules is created from the

⁴ SKOS vocabulary, <http://www.w3.org/2009/08/skos-reference/skos.html>

⁵ Apache Jena, <http://jena.apache.org/>

⁶ Drools – Business logic integration platform, <http://www.jboss.org/drools/>

Table 1. Evaluation between the 2 Panacea reasoning approaches and GalenOWL

	Panacea-Jena	Panacea-Drools	GalenOWL
Initialization time	32.0 s	34.7 s	148 s
Memory consumption	169 MB	280 MB	649 MB
Query response time	47 ms	5 ms	16 ms

initial rule base that match the patient’s sex and age group, and a fine selection phase where rules from \mathcal{A} that contain at least one of the patient’s data, i.e. a *skos* term, in their premises are singled out and a fine set $\mathcal{R} \subseteq \mathcal{A}$ is created from them. \mathcal{R} is loaded in Jena and can contain as little as 50 rules.

3 Evaluation and Conclusion

For evaluating the framework, a comparison was made between the two approaches for the rule-based reasoning and GalenOWL. The comparison was focused on the usability of the framework in a production environment. Three parameters were measured: initialization time, memory consumption after initialization, and query response time, i.e. the time needed to retrieve the recommendations. Results are shown in Table 1, where values for GalenOWL are from [2]. The results show that both Panacea approaches exhibit satisfactory results on all three parameters with Jena appearing to consume less memory as the rule base is loaded during query time and with Drools being optimized to handle large rule bases efficiently.

The paper presented Panacea, a framework for semantic-enabled drug recommendations discovery. The framework utilizes a layered reasoning approach where the medical ontology and the patient data instances are fed to a RDF(S) reasoner and drug recommendations are generated using the second, rule-based reasoning layer where any rule engine can be employed. Concerning future work, the possibility to add probabilistic reasoning will be investigated and an evaluation of the framework on real world data will be carried out.

Acknowledgments This work has been supported by the national project “Panacea”, funded by GSRT Hellas under the “Support for SMEs” programme.

References

1. Adnan, M., Warren, J., Orr, M.: Ontology based semantic recommendations for discharge summary medication information for patients. In: Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on. (oct. 2010) 456–461
2. Doulaverakis, C., Nikolaidis, G., Kleontas, A., Kompatsiaris, I.: GalenOWL: Ontology based drug recommendations discovery. *Journal of Biomedical Semantics* **3**(14) (2012)