Development and Population of an Elaborate Formal Ontology for Clinical Practice Knowledge Representation

David Mendes[1], Irene Pimenta Rodrigues[1] and Carlos Fernandes Baeta[2]
[1]*Departamento de Informática, Universidade de Évora, 7004-516 Évora, Portugal*
[2]*Departmento de Medicina, Hospital José Maria Grande, Portalegre, Portugal*

Keywords: OGCP, OGMS, CPR, Ontological Realism, SOAP, Clinical Practice Knowledge, OWL.

Abstract: We introduce the *Ontology for General Clinical Practice* **(OGCP)** for better knowledge representation support in the Clinical Practice domain. We followed the established OBO Foundry principles to leverage the ontological relations that might be present in the ontology axioms we harvest from clinical reports text segments. In accordance to the *Ontological Realism* principles we expect the reasoning inferred from the ontological relations to render more acceptable consequences then logical relations alone. We enhance the *Ontology for General Medical Science* ***(OGMS)*** *with the Computer-Based Patient Record Ontology* ***(CPR)*** *structure and propose knowledge base creation/enhancing automatically extracting from clinical reports written in the, well known to the medical community, SOAP format. Reasoning over the resulting* ***(OGCP)*** *knowledge base with novel parallel algorithms that appeared recently in literature is presented. We finally propose* ***Controlled Natural Language*** *justifications of the inferred knowledge intending to achieve wider acceptance among clinicians.*

1 MOTIVATION AND RESEARCH QUESTIONS

Originally our research intention was the development of personal CDS[1] tools to help the healthcare professionals in scarce resource countries like most in Africa and Asia. After evaluating the State-of-the-Art presented ahead we found that relevant work is yet to be done in the KR[2] area regarding the Clinical Practice domain. We believe that some developments that have been achieved recently motivate us to incorporate our expertise in NLP[3] into effective ontology population. Our main intention is to be able to *automatically produce clinical practice knowledge bases* extracting from healthcare reports text.

Research Questions. Ontologies in the sub-domain of *Clinical Medicine*[4] are lacking some thorough study. These can be stated as current problems for the effectiveness of using them as knowledge support for *clinical reasoning*. Problems found in current ontologies and enumerated in literature (Hoehndorf et al., 2011) that lead to reasoning hurdles are:

- *Lack of adequate modularization* to achieve the minimum amount of implicit differentiation among primitive concepts.
- Inadequate *clear separation of digital entities from the reality they represent*.
- Inability to avoid the *knowledge acquisition bottleneck* (Wong et al., 2012) in order to speed start any automatic enrichment.

In our work we try to overcome the different issues identified by the several experts in (Brochhausen et al., 2011). In order to maximize the reasoning capabilities based in our extended OGMS(OGMS, 2010) ontology, different considerations in the referred work by Brochhausen et al. were taken into good account. We complemented the **OGMS** ontology with the **CPR** into what we call the **OGCP** that is intended to be a more supportive structure for representation of clinical practice while, at the same time, embodies a formal medical theory of disease and healthcare.

[1]Clinical Decision Support
[2]Knowledge Representation
[3]Natural Language Processing
[4]The study of disease by direct examination of the living patient

Mendes D., Rodrigues I. and Fernandes Baeta C..
Development and Population of an Elaborate Formal Ontology for Clinical Practice Knowledge Representation.
DOI: 10.5220/0004548602860292
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2013), pages 286-292
ISBN: 978-989-8565-81-5

2 STATE OF THE ART

Our previous work (Mendes and Rodrigues, 2012b), where an exhaustive state of the art is presented, is focusing mainly over the CPR ontology (W3C, 2009) but we are now targeting an extended OGMS because it is more promising as suitable for representation of a disease theory and model enhancing the corresponding reasoning capabilities. We base our work in (Ogbuji, 2011) for what matters about the foundational principles of structuring meaningful knowledge representation as a framework for clinical reasoning. We consider the above mentioned (Smith and Ceusters, 2010) for Ontological Realism approach. We used the excellent recent review by Wong (Wong et al., 2012) for updated state of the art problems in knowledge acquisition from text. We explore the recent achievements in controlled natural language generation presented in (Kaljurand, 2010) with the distributed processing possibilities suggested in (Kazakov et al., 2011) for consequence based axiom inferring introduced in (Simancik et al., 2011).

3 APPROACH

We illustrate the relations between open questions and the current line of work to illustrate the used approach:

Ontological Realism and Relations. The methodology to avoid mistakes that cannot be detected by logical formalisms alone is the formal use of Ontological Realism. We highlight the reasoning power that formal ontological relations provide to a carefully crafted ontology given the higher semantic level that these relations comprise (Smith and Ceusters, 2010). The formalization of *Ontological Relations* has been advocated for many years and it succeeded in the development of *"relations that obtain between entities in reality, independently of our ways of gaining knowledge about such entities"* (Smith et al., 2005).

OGCP as Suitable Support for Clinical Practice Knowledge. It is an ontology of entities involved in a clinical encounter. OGCP includes very general terms that are used across medical disciplines, including: 'disease', 'disorder', 'disease course', 'diagnosis', 'patient', and 'healthcare provider'. OGCP uses the Basic Formal Ontology (BFO) (BFO, 2012) as an upper-level ontology. OGCP provides a formal theory of disease and treatment. This theory is implemented using OWL-DL and is available in OWL.

SNOMED CT as the Primary Terminology Aggregation. Our effort will take advantage of the breadth of coverage of SNOMED CT in our domain of interest(Smith and Brochhausen, 2010).
It has an underlying description logic ($\mathcal{EL}$ family). $\mathcal{EL}$ family has shown to be suitable for medical terminology processing and subsequently, $\mathcal{ELHR}+$ is the performance target of many modern classifiers including those based in consequence driven reasoning capable of classifying SNOMED CT in practical and acceptable processing times with recent proposed extension for concurrent processing (Simancik et al., 2011) that benefits of current advances in *BigData* cluster processing.

OGCP abiding to OBO Foundry. The upper-level ontologies that support OGCP are introduced in what regards the ontological relations that can and shall be used: *BFO* (IFOMIS, 2004) is strictly focused on the task of providing a genuine upper ontology which can be used in support of domain ontologies developed for scientific research within the framework of the OBO Foundry; *FMA*; *RO*(RO, 2012) and *AIAO*.

Clinical Ontology Fine Tuning. Starting from the ontology alignment in the following figure.

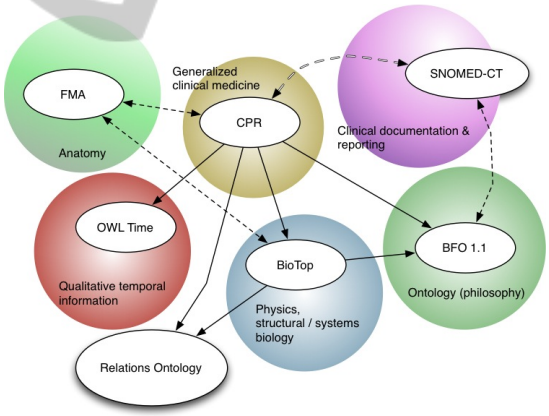


Figure 1: Ontology Alignment.

We made an effort of trimming and pruning of the OGMS and CPR complementing in accordance to our team of cardiologists to better accommodate their needs expressed in the reports we sampled. That included some "gardening" to include: **SO** the Symptom Ontology, **VSO** the Vital Signs Ontology and others all of them accord to OBO Foundry principles. In order to align the clinical concepts in the various ontologies present, an effort was needed to amalgamate them according to a sound theory of disease and that's why we incorporate the **DO** that was expressly built with this purpose in mind (Schriml et al., 2012). The

Disease Ontology is a community driven, open source ontology that is designed to link disparate datasets through disease concepts. It's provided a computable structure of inheritable, environmental and infectious origins of human disease to facilitate the connection of genetic data, clinical data, and symptoms through the lens of human disease (Wiki-DO, 2012). The DO semantically integrates disease and medical vocabularies through extensive cross mapping and integration of **MeSH**, **ICD**, **NCI Thesaurus**, **SNOMED CT** and **OMIM** (OMIM, 2012) disease-specific terms and identifiers. It represents a comprehensive knowledge base of 8043 inherited, developmental and acquired human diseases.

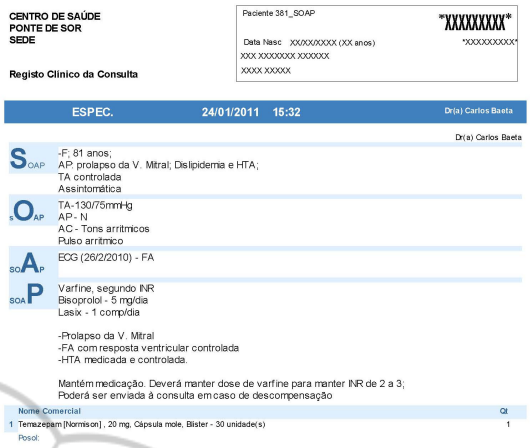
Ontology Learning From Text. The problem of acquiring the knowledge necessary for ontology population known as the *"Knowledge Acquisition Bottleneck"* is a challenging (Wong et al., 2012) issue that remains one of the main barriers for automated acquisition and we tried to circumvent it by using a progressive tutored learning approach. We start from semi-structured text and use the semi-automated translation tasks to generate a controlled *domain specific vocabulary* on which further acquisition tasks build upon (Mendes and Rodrigues, 2012b) minimizing ambiguity and redundancy for better reasoning capabilities. When instantiating individuals (populate) formal heavyweight ontologies like the OGCP we do not normally intend to enrich the ontology but instead turn them from theoretical models of the domain into *reasoning able* knowledge bases.

4 RESEARCH METHODOLOGY

As reviewed in (Wong et al., 2012) the state-of-the-Art for acquisition from Clinical Text has enjoyed strong developments in recent years. We are diving into extraction from free text present in most interfaces used by clinicians. So far we are elaborating in Cardiovascular related healthcare.

4.1 Text Ontological Annotation

To convert the source texts (clinical note, release report, exam report, or others) the process is a transforming sequence that involves several sequential steps. This transformation renders ultimately an OGCP instance. These tasks can be done manually or automated. Those steps workflow can be configured declaratively using the software architecture shown in section 5.1. There are steps involved that consist of:


CENTRO DE SAÚDE
PONTE DE SOR
SEDE

Registo Clínico da Consulta

Paciente 381_SOAP *XXXXXXXXX*
Data Nasc XXXX/XXXX (XX anos) *XXXXXXXXX*
XXX XXXXXXX XXXXXX
XXXX XXXXX

ESPEC. 24/01/2011 15:32 Dr(a) Carlos Baeta

Dr(a) Carlos Baeta

S -F, 81 anos;
AP prolapso da V. Mitral; Dislipidemia e HTA;
TA controlada
Assintomática

O TA-130/75mmHg
AP - N
AC - Tons arrítmicos
Pulso arrítmico

A ECG (26/2/2010) - FA

P Varfine, segundo INR
Bisoprolol - 5 mg/dia
Lasix - 1 comp/dia

-Prolapso da V. Mitral
-FA com resposta ventricular controlada
-HTA medicada e controlada.

Mantém medicação. Deverá manter dose de varfine para manter INR de 2 a 3;
Poderá ser enviada à consulta em caso de descompensação

	Nome Comercial	Qt
1	Temazepam [Normison] , 20 mg, Cápsula mole, Blister - 30 unidade(s)	1

Posol:

Figure 2: SOAP Report Sample.

- PDF to raw text or to structured (XML) converting for adequate documents cleansing. For instance the graphical presentation of Vital Signs that are originally rendered in the respective report has to be deleted from the document for easier terms processing and the tables with values must be structured accordingly for the annotators to behave properly. Initially there is a proof of concept that involves manually cleaning the original reports
- Manual translation (that is indispensable for the translator tutoring as shown in 4.2) with the precise clinicians validation of their jargon adequately translated into English,
- Annotation, either manually using the Web interface of any of the services that we introduce in 5.2, or automatically through the Web Services available
- Filtering the concepts from the annotated text to insert in OGCP instances

Given the array of available Web Services that can semantically annotate bio-medical concepts in English that are presented ahead in section 5.2, we chose to use an evolutionary approach for use of the BioPortal annotator (Noy et al., 2009). We mean by evolutionary approach the fact that we first use the annotator after manual pre-processing and then a more automatic workflow.

4.2 Using Automated Translation for Concept Unification

We can take advantage of the fact that we have to translate from jargon to English to customize the Google translator toolkit[5] with our own Translation

[5]https://translate.google.com/toolkit

Memories and Glossaries. Let us introduce some demonstrative examples taken from the sample document gently provided by Dr. Carlos Baeta and properly de-identified presented in the figure "SOAP Report Sample" above.

We will, in the process of using the Google toolkit, create Translation Memories with the identified personal acronyms like:

- AP (Antecedentes Pessoais) into Personal History
- HTA (Hiper Tensão Arterial) into High Blood Pressure
- FA (Fibrilhação Auricular) into Atrial Fibrilation
- V. Mitral (Válvula Mitral) into Mitral Valve

Some which are acronyms that can be given the suitable translated concept like:

- ECG (Electro Cardio Grama) into Electro Cardio Gram

or those that are even English acronyms:

- INR (International Normalized Ratio) into International Normalized Ratio

Included in this sample are notorious some more complex problems that are not related to the translation itself but with some other problems like the time spanning of concepts like “1 comp/dia” which is adequately translated to “1 tablet per day” using the defined Translation Memory but has to be posteriorly exactly characterized as time delimited occurring process.

4.3 SOAP Report

This report depicts a clinical encounter in a semi-structured way. As seen previously in the figure in this section we find sections that can be associated with

Symptoms, the subjective section S where we extract directly to `ogcp:symptom-recording`.

Signs, the objective section O that are `ogcp:sign-recording` that we take as generator for `ogcp:clinical-findings`.

Actions, the analysis section A which are the ogcp:clinical-investigation-act whose outputs can be `ogcp:clinical-artifact` to investigate things that can be `ogcp:isConsequenceOf` any of `ogcp:physiological-process` or `ogcp:pathological-process`

and finally

Plan, the plan section P where the therapeutic acts can be extracted with all the timing, posology and prescriptions registered in a particular clinical encounter.

We find then that if a **sufficient amount of clinical reports** are fed into the knowledge base it will ultimately build a sound picture of a clinical practice. For the inferred axioms to be believable by the community and thus usable as a Clinical Practice Supporting Tool that generated knowledge has to be clearly, although basically, explained to the user.

4.4 The Complete Acquisition Workflow Picture

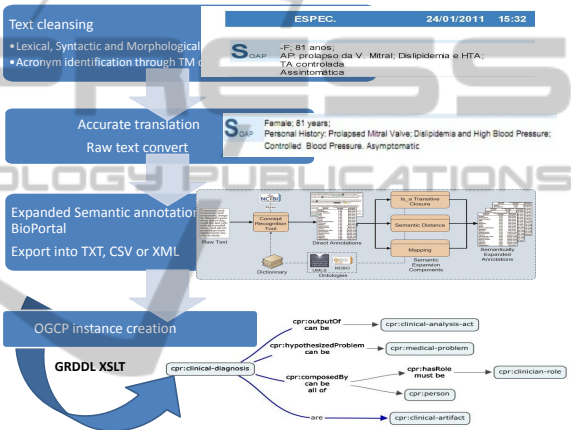


Figure 3: Acquisition Flowchart.

The flowchart that depicts graphically the acquisition from the source texts in Portuguese to the creation of the appropriate OGCP instance is shown in figure "Acquisition Flowchart"

5 RESULTS

5.1 Software Implementation

We have presented the full conception of an architecture in (Mendes and Rodrigues, 2012a). We are proposing an extensible Enterprise Architecture based in a ESB which we call CP-ESP (Clinical Practice - Enterprise Service Bus). This is a common rail where messaging can flow using a subscription model that enables the communication to be detached from any two particular services but instead be available on-request by one and served by another in a loosely coupled way. The ESB can then intervene in message exchange and overwrite standard rules for service execution. The case of an intervention here is

the ability to filter and redirect invocations to the appropriate NLP task processors depending on the source being labeled with the status of the load it carries. The REST philosophy is suggested in our proposal as the best way of implementing a Service Oriented Architecture that serves as the communication underlying structure of our system. REST endpoints are available for the generality of our needs. The available Web Services can render responses in highly-structured forms like JSON or in any of the standardized mime types that can be handled by the filtering and enqueuing capabilities of any configurable available ESB like those based in Apache ServiceMix (http://servicemix.apache.org/home.html) or Mule (http://www.mulesoft.org/) for instance. They can be configured to compose a complete pipeline very easily:

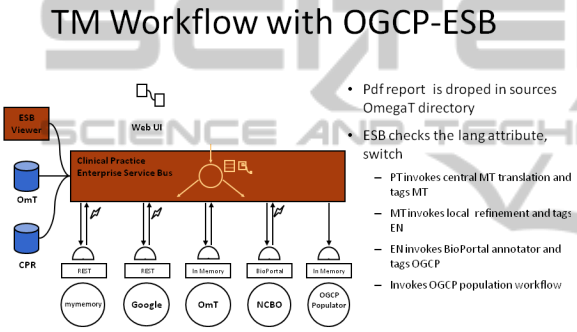


Figure 4: Instance Translation and Creation.

In the picture is shown the possibility of defining services (daemons) that monitor the presence of new reports and appropriately apply the needed transformations according to their status and content until a translated document is delivered to the adequate endpoint for annotation. All of this can even be done with due care about scalability, availability and all problems associated with a modern state-of-the-art software architecture as presented in (Mendes and Rodrigues, 2012a).

In the picture is shown the possibility of defining services (daemons) that monitor the presence of new reports and appropriately apply the needed transformations according to their status and content until a translated document is delivered to the adequate end-point for annotation. All of this can even be done with due care about scalability, availability and all problems associated with a modern state-of-the-art software architecture as presented in (Mendes and Rodrigues, 2012a) Building over the suggested infrastructure the systems are rather composed as opposed to monolithically built and so manifest high capabilities of plug-and-play configuration allowing for interchangeable providers (as Web Services), Reference Ontologies (Feeders), and target ontologies. Having the foundations available with the right weapons provided one has to take a practical approach to the development of a target system using, in our case, the Java best-practices for pragmatic development that include a number of Patterns as in JEE (Java Enterprise Edition) or the pragmatic approaches developed in such successful projects as Spring (http://www.springsource.com/).

5.2 OGCP-ESB

We use, through BioPortal and ODIE, algorithms for Named Entity Recognition, Co-reference resolution, concept discovery, discourse reasoning and attribute value extraction. A local version gives the possibility of developing extensions to the algorithms provided in the base offering allowing, for instance, targeting different languages in the NLP tasks. The Web Services provided by BioPortal or ODIE can be locally extended and refined for different sources and are provided as one of the project deliverables.

5.3 NLP Pragmatics and Discourse Controller

The text for any particular encounter (actually for any Clinical Episode) may be collected in the form suitable for processing into the Ontology framework using some **NLP pragmatics**. Populating the *OGCP* the "Clinical Picture" is completed and thus our KB is available for validation and further logical inferencing. The semantic representation is done using pragmatic interpretation as defined in our fellow researcher at CENTRIA Dora Melo's article (Melo et al., 2012). The enrichment process must always maintain the entailments provided by the base (gold-standard) ontologies and so can never lead to inconsistency. We use a round-trip, debug and repair, building method to populate/enhance the OGCP then. For any new instance the validation is performed and new possible inferred facts generated if consistency is yet valid. These new facts are candidates for NLP justifications generation. The main objective of the system is to provide accurate answers to questions posed by users and, in our proposal these answers are clinically valid because the generation method guarantees that. QA is, however, only one of the interesting features of our work that is enhanced by the adequate justification to be evidently useful for practitioners. To develop justifications from DL[6] arguments inferred from consequence based reasoners (Kazakov, 2009;

[6]Description Logic

Kazakov et al., 2011) we based our work in (Bail et al., 2011) to study and compare the justificatory structure to those present in the NCBO BioPortal addressed in the mentioned article. The results so far are in the realm of **'ontology verbalization'**, the generated explanations are still in a **controlled natural language (CNL)** fashion. The obtained results seem to be adequate enough for the users to find them believable and thus the justifications stand in our controlled clinical setting. We use the verbalization tooling (Kaljurand, 2010) to present the justifications in an acceptable manner. The foundational techniques were introduced in (Kaljurand and Fuchs, 2007). For the verbalization to function properly all the restrictions of content are guaranteed in the process of ontology (Knowledge base) enrichment from SOAP reports. For instance, all names are English words and individuals are singular proper names (preferably capitalized) named classes are denoted by singular countable nouns and (object) properties by transitive verbs in their lemma form (i.e. infinitive form) (Kaljurand and Fuchs, 2007). The decision of what inferred knowledge is then presented with its justifications to the user is a task handled by the DC[7] using the developed pragmatics introduced in the above referred article (Melo et al., 2012). In our proposed methodology we make use os the DC to align the most relevant clinical terms into a acceptable CNL document and the process may be graphically presented as:

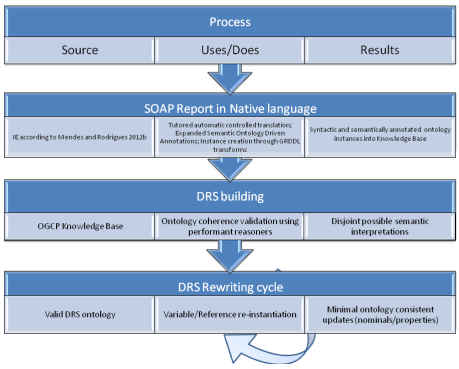


Figure 5: Discourse Controller Flow.

6 CONCLUSIONS

We propose an **OGMS** extension using the adequate ***ontological realism*** approaches and incorporating the **CPR** and its upper level ontologies as framework for an $\mathcal{EL}$ reasoning workhorse. We present our efforts for knowledge base population from semi-structured clinical text reports and discuss the underlying problem of automatic instance creation from them into the proposed knowledge representation structure. Currently, after having passed an internal process of validating the proposed framework *(Ontology structure+Population mechanism)* we are assessing its acceptance in a wider regional level by exposing a group of related cardiologists to the works done so far and receiving their feedback in a formal and traceable manner. We are developing a knowledge representation infrastructure enabling the usage of highly optimized distributed consequence based reasoners that are referred in literature only in 2011. With these very recent developments it's finally possible to validate the enormous knowledge bases that are created by automatically populating the proposed ontology *OGCP* that relies on extensive, and very solid, foundations like *SNOMED-CT* and *FMA* among others. Logical inferencing and clinical facts entailment that is possible through this capability is an interesting contribution to the application of Artificial Intelligence to healthcare. We introduce clinical decision support systems (*CDSS*) that are based on such a breakthrough technique. We further argue that it is imperative, for the broad acceptance of these tooling by the medical community, that their inferences are justified using controlled natural language and adequate terminology.

7 FUTURE WORK

We intend to deliver *OGCP* in the NCBO BioPortal as soon as it reaches a minimum of usability as KR tool for cardiology which we intend to happen later this year. We are currently in the evaluation process mentioned above and, depending on the level of usability/acceptance, will evolve to different clinical specialties to demonstrate its flexibility. Shall all the efforts prove to be worthy we will try to extend it to wider levels following a community development process based in some OSS (Open Source Software) repository like *Google Code* for example. It will all be summed up in the PhD thesis of the first author to be presented late 2013.

ACKNOWLEDGEMENTS

We acknowledge IIFA - Instituto de Investigação e Formação Avançada of Universidade de Évora for the Bento de Jesus Caraça scholarship that is awarded to the first author and CENTRIA Center for Research in Artificial Intelligence of Faculdade de Ciências e Tec-

[7]Discourse Controller

nologia of Universidade Nova de Lisboa for its continuous financial support of our work.

REFERENCES

Bail, S., Horridge, M., Parsia, B., and Sattler, U. (2011). The justificatory structure of the ncbo bioportal ontologies. In Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N. F., and Blomqvist, E., editors, *International Semantic Web Conference (1)*, volume 7031 of *Lecture Notes in Computer Science*, page 67Ã¢âĆňâĂIJ82. Springer.

BFO (2012). Basic formal ontology 2.0.

Brochhausen, M., Burgun, A., Ceusters, W., Hasman, A., Leong, T., Musen, M., Oliveira, J., Peleg, M., Rector, A., Schulz, S., et al. (2011). Discussion of" biomedical ontologies: Toward scientific debate". *Methods of information in medicine*, 50(3):217.

Hoehndorf, R., Dumontier, M., Oellrich, A., Rebholz-Schuhmann, D., Schofield, P., and Gkoutos, G. (2011). Interoperability between biomedical ontologies through relation expansion, upper-level ontologies and automatic reasoning. *PloS one*, 6(7):e22006.

IFOMIS (2004). Basic formal ontology.

Kaljurand, K. (2010). Owl verbalizer: making machine-readable knowledge also human-readable. Available Online: https://code.google.com/p/owlverbalizer/.

Kaljurand, K. and Fuchs, N. E. (2007). Verbalizing owl in attempto controlled english. *Proceedings of OWLED07*.

Kazakov, Y. (2009). Consequence-Driven Reasoning for Horn SHIQ Ontologies. In Boutilier, C., editor, *IJCAI*, pages 2040–2045. IJCAI Distinguished Paper Award Winner.

Kazakov, Y., Krötzsch, M., and Simancík, F. (2011). Concurrent classification of el ontologies. In *Proceedings of the 10th international conference on The semantic web - Volume Part I*, ISWC'11, pages 305–320, Berlin, Heidelberg. Springer-Verlag.

Melo, D., Rodrigues, I., and Nogueira, V. (2012). Work out the semantic web search: The cooperative way. *Advances in Artificial Intelligence*, 2012.

Mendes, D. and Rodrigues, I. P. (2012a). A Semantic Web pragmatic approach to develop Clinical Ontologies, and thus Semantic Interoperability, based in HL7 v2.xml messaging. In *Information Systems and Technologies for Enhancing Health and Social Care*. IGI Global.

Mendes, D. and Rodrigues, I. P. (2012b). Advances to semantic interoperability through cpr ontology extracting from soap framework reports. *electronic Journal of Health Informatics*.

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., and Musen, M. a. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, 37(Web Server issue):W170–3.

Ogbuji, C. (2011). A Framework Ontology for Computer-Based Patient Record Systems. In *Proceedings of the ICBO: International Conference on Biomedical Ontology*, pages 217–223, Buffalo, NY, USA.

OGMS (2010). Ontology for general medical science.

OMIM (2012). Omim, online mendelian inheritance in man. Available Online: http://www.ncbi.nlm.nih.gov/omim.

RO, O. F. (2012). Relations ontology.

Schriml, L., Arze, C., Nadendla, S., Chang, Y., Mazaitis, M., Felix, V., Feng, G., and Kibbe, W. (2012). Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Research*, 40(D1):D940–D946.

Simancik, F., Kazakov, Y., and Horrocks, I. (2011). Consequence-Based Reasoning beyond Horn Ontologies. In Walsh, T., editor, *IJCAI*, pages 1093–1098. IJCAI/AAAI.

Smith, B. and Brochhausen, M. (2010). Putting Biomedical Ontologies to Work. *Methods Inf Med*, 49.

Smith, B. and Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied ontology*, 5(3-4):139–188.

Smith, B., Ceusters, W., Klagges, B., KÃűhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L., and Rosse, C. (2005). Relations in biomedical ontologies. *Genome Biol*, 6(5):R46.

W3C (2009). Computer-based patient record ontology. http://code.google.com/p/cprontology. http://code.google.com/p/cprontology.

Wiki-DO (2012). Disease ontology wiki page. Available Online: http://do-wiki.nubic.northwestern.edu/do-wiki/index.php/Main_Page.

Wong, W., Liu, W., and Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM Comput. Surv.*, 44(4):20:1–20:36.