A Description Logic for InferenceNet.Br

Wellington Franco, Thiago Alves, Henrique Viana, and João Alcântara

Departamento de Computação, Universidade Federal do Ceará, P.O. Box 12166,
Fortaleza, CE, Brasil 60455-760
{jwellingtonfranco,thiagoalves,henriqueviana,jnando}@lia.ufc.br

Abstract. The InferenceNet.Br is a new linguistic resource for Portuguese lan-
guage knowledge bases, which follows some principles of the WordNet, Con-
ceptNet and FrameNet. Unlike these other linguistic resources, InferenceNet.Br
allows representation of more expressive relationships between concepts, permit-
ting to characterize premises or conclusions of these relationships. However, the
integration between InferenceNet.Br with other linguistic resources is not possi-
bile directly. In order to settle this problem, in this paper we present a mapping
of the knowledge bases of InferenceNet.Br to the Description Logic
DL-Lite_A,
which is a fragment of OWL. As we know, OWL is a standard language designed
to facilitate machine interpretability of Web content, through explicitly represen-
tation of the meaning of concepts in vocabularies and relationships between them.
By doing this, we provide a connection between InferenceNet.Br and resources
as WordNet and ConceptNet, since there are some efforts to link them to OWL.

Keywords: Linguistic Resource, InferenceNet.Br, Description Logic, OWL.

1 Introduction

Natural Language Processing (NLP) is a field of computer science and linguistics con-
cerned with technics to construct computational systems which treat different levels of
meanings and uses of natural languages. Roughly speaking, NLP systems are conceived
to enable computers to manipulate linguistic signs reasonably in order to take decisions,
extract and retrieve information, summarize texts, and solve other tasks involving the
understanding of sentences and texts in natural language.

In Mitkov [21], it was stated that the reasoners of current NLP systems employ
predominantly syntactic approaches, facing difficulties in capturing certain kind of
knowledge, mainly because of the lack of linguistic resources that support a com-
plete understanding of concepts and sentences. The usual reasoners define taxonomies
or ontologies through objects of determined domain, classes of objects and relation-
ships between objects, but do not represent their practical content. For instance, when
we take into consideration the word chair, we can either conclude that it is made of a
specific material, as wood for example, or we can conclude that it has some practical
utilities, as a seat or even a weapon.

An alternative applied to improve the quality of inferences in NLP systems is to
resort to lexical-semantic base systems, whose aim is to add contextual elements that
enhance the ability of these systems. The three principal bases used in semantic sys-
tems and applications of NLP are the WordNet [19], the FrameNet [6] and the Concept-
Net [15]. WordNet provides lexical-semantic resources for the Portuguese language;
however it expresses only basic semantic hierarchy relationships (hyperonymy and hy-
ponymy), inclusions (holonymy and meronymy), equivalences (synonymy) or opposi-
tions (antonymy), and it is not available for large-scale use. FrameNet is based on the
theory of frames proposed by Minsky [20]. A frame is a conceptual hierarchical struc-
ture defining a situation, object or event by participants and their relationships. Con-
ceptNet is a commonsense knowledge base in Portuguese language, which creates a net
of relationships between concepts, such that these relationships may express causality,
functionality, location, time, etc.

The Semantic Inferencialism Model (SIM) [22,23] is another lexical-semantic base
with good results. It consists of a new model for treatment of pragmatic-semantic level
of natural languages, adding new information in relationships and allowing richer in-
ferences. SIM comprises four components: Conceptual Base, Sentence Patterns Base,
Rule Base for Practical Reasoning and an algorithm to deduce information from these
knowledge bases. To represent Conceptual and Sentence Patterns Bases it is used the
Portuguese linguistic resource InferenceNet.Br [24] and the Rule Base for Practical
Reasoning is represented as a knowledge base in logic programming [16].

ConceptNet, the InferenceNet.Br is a commonsense knowledge base which can
represent many different types of knowledge, as relationships between concepts, ex-
pressed as simple phrases of natural language. This resource can be accessed by the
portal www.inferencenet.org, where its Conceptual Base contains around 180.000 rela-
tionships between concepts and 700.000 relationships of pragmatical content, and the
Sentence Patterns Base contains 5963 sentence patterns through 1061 relationships.

However, the InferenceNet.Br is not directly related to any other resources known,
as WordNet, FrameNet and ConceptNet. Following the works [12,14], which converts
WordNet and ConceptNet databases in Web Ontology Language (OWL) [13], the
standard language to represent ontologies in the Semantic Web, in this paper we will
translate InferenceNet.Br into a Description Logic namely DL-Lite_4 [7]: a logic of the
DL-Lite family that can be expressed in OWL-Lite [13,18] and that is expressive enough
to represent InferenceNet.Br knowledge bases. By doing this translation, we can estab-
lish a connection between InferenceNet.Br with databases of WordNet and ConceptNet.

This paper is organized as follows: Section 2 describes the general aspects of In-
ferenceNet.Br and explains the component and main characteristics of the resource. In
Section 3 we describe the Description Logic DL-Lite_4 and we show how to represent
the content of Conceptual and Sentence Patterns Bases using this logic. Finally, we
conclude the paper in Section 4.

2 InferenceNet.Br

One motivation for introducing a new linguistic resource is the lack of linguistic
resources with large scale semantic knowledge for the Portuguese language. Lexical-
semantic bases in Portuguese as WordNet.Pt [10], WordNet.Br [9], TeP 2.0 [17], PA-
PEL [11], FrameNet Brasil [25] or OMCS-Br [1] are restricted to specific domains, i.e.,
they depend of a text corpus, a dictionary or a thesaurus [26]. Another motivation is the
non-existence of a linguistic resource with inferentialist semantic knowledge, either for
Portuguese or other languages [22,23]. InferenceNet.Br contains two knowledge bases