

# Proceedings of the Workshop on Natural Language Processing and Automated Reasoning (NLPAR)

collocated with 12th International Conference  
on Logic Programming and Nonmonotonic Reasoning (LPNMR)

September 15, 2013  
A Corunna, Spain

Organizing Committee:

Chitta Baral (Arizona State University, USA)

Peter Schüller (Sabanci University, Turkey)

## Invited Papers

*Some Recent Advances in Answer Set Programming*

Marcello Balduccini

Answer Set Programming (ASP) is a logical language for knowledge representation and reasoning that combines a non-monotonic nature, strong theoretical foundations, an intuitive semantics, and substantial expressive power. The language has been successfully used for modeling a number of very diverse domains and for capturing key reasoning tasks such as planning, diagnostics, learning and scheduling. All of this makes ASP a prime candidate for use in the sophisticated knowledge representation and reasoning tasks involved in Natural Language Processing (NLP). In this talk I will give an overview of some of my recent work on ASP that I believe may find application in the context of NLP.

*Three things NLP should consider to enable automated reasoning*  
Vinay Chaudhri, Nikhil Dinesh and Daniela Incezan

Our work is driven by the hypothesis that for a program to answer questions, explain the answers, and engage in a dialog just like a human does, it must have an explicit representation of knowledge. Such explicit representations occur naturally in many situations such as engineering designs created by engineers, a software requirement created in unified modeling language or a process flow diagram for a manufacturing process. Automated approaches based on natural language processing have progressed on tasks such as named entity recognition, fact extraction and relation learning. Use of automated methods can be problematic in situations where the conceptual distinctions used by humans for reasoning are not directly expressed in natural language or when the representation must be used to drive a high fidelity simulation. In this paper, we report on our effort to systematically curate a knowledge base for substantial fraction of text in a biology textbook [26]. While this experience and the process is interesting on its own, three aspects can be especially instructive for future development of knowledge bases by both manual and automatic methods: (1) Consider imposing a simplifying abstract structure on natural language sentences so that the surface form is closer to the target logical form to be extracted. (2) Adopt an upper ontology that is strongly motivated and influenced by natural language. (3) Develop a set of guidelines that captures how the conceptual distinctions in the ontology may be realized in natural language. Since the representation created by this process has been quite effective for answering questions and producing explanations, it gives a concrete target for what information should be extracted by the automated methods.

## Regular Papers

*BOEMIE: Reasoning-based Information Extraction*

Georgios Petasis, Ralf Möller and Vangelis Karkaletsis

This paper presents a novel approach for exploiting an ontology in an ontology-based information extraction system, which substitutes part of the extraction process with reasoning, guided by a set of automatically acquired rules.

## *Qualitative Analysis of Contemporary Urdu Machine Translation Systems*

Asad Abdul Malik and Asad Habib

The diversity in source and target languages coupled with source language ambiguity makes Machine Translation (MT) an exceptionally hard problem. The highly information intensive corpus based MT leads the MT research field today, with Example Based MT and Statistical MT representing two dissimilar frameworks in the data-driven paradigm. Example Based MT is another approach that involves matching of examples from large amount of training data followed by adaptation and re-combination. Urdu MT is still in its infancy due to nominal availability of required data and computational resources. This paper provides a detailed survey of the aforementioned contemporary MT techniques and reports findings based on qualitative analysis with some quantitative BLEU metric quantitative results. Strengths and weaknesses of each technique have been brought to surface through special focus and discussion on examples from Urdu language. The paper concludes with proposal of future directions for research in Urdu machine translation.

## *The NL2KR System*

Chitta Baral, Juraj Dzifcak, Kanchan Kumbhare and Nguyen Vo

In this paper we will describe the NL2KR system that translates natural language sentences to a targeted knowledge representation formalism. The system starts with an initial lexicon and learns meaning of new words from a given set of examples of sentences and their translations. We will describe the first release of our system with several examples.

## *Recognizing Implicit Discourse Relations through Abductive Reasoning with Large-scale Lexical Knowledge*

Jun Sugiura, Naoya Inoue and Kentaro Inui

Discourse relation recognition is the task of identifying the semantic relationships between textual units. Conventional approaches to discourse relation recognition exploit surface information and syntactic information as machine learning features. However, the performance of these models is severely limited for implicit discourse relation recognition. In this paper, we propose an abductive theorem proving (ATP) approach for implicit discourse relation recognition. The contribution of this paper is that we give a detailed discussion of an ATP-based discourse relation recognition model with open-domain web texts.

## *A default inference rule operating internally to the grammar devices*

Christophe Onambélé Manga

Minimalist Grammars (MG) are viewed as a resource consuming system where syntactic operations are triggered when a positive form of a feature matches with its negative form. But a problem arises when a feature lacks a positive/negative value. For the latter case, we introduce a default inference rule in order to account for the underspecification of the feature in a lexical entry.