

Computational Linguistics and Intellectual Technologies:  
Proceedings of the International Conference “Dialogue 2017”

Moscow, May 31—June 3, 2017

## **OPEN KNOWLEDGE REPRESENTATION FOR TEXTUAL INFORMATION**

**Ido Dagan** (ido.k.dagan@gmail.com)

Bar-Ilan University, Israel

How can we capture the information expressed in multiple texts? And how can we allow people, as well as computer applications, to easily explore it? When comparing textual knowledge to formal knowledge representation (KR) paradigms, two prominent differences arise. First, typical KR paradigms rely on pre-specified vocabularies, which are limited in their scope, while natural language is inherently open. Second, in a formal knowledge base each fact is encoded in a single canonical manner, while in multiple texts facts may be repeated with some redundant, complementary and even contradictory information.

In this talk, I will outline a new research direction, which we term Open Knowledge Representation (OKR), that aims to represent textual information in a consolidated structured manner, based on the available natural language vocabulary and structure. I will describe our first specification for Open Knowledge Graphs, motivated by a use case of representing multiple tweets describing an event, for which we have created a medium-scale annotated dataset. Our proposed structure merges co-referring individual proposition extractions, created in an Open-IE flavor, into a representation of consolidated entities, predicates and statements, inspired by traditional knowledge graphs. Information redundancy is further modeled via entailment relations. I will also illustrate the potential application of our open knowledge graphs for text exploration and point at possible directions in which the OKR paradigm might evolve.