

Entailment graphs for text exploration

Ido Dagan[†] **Bernardo Magnini**[‡]

[†] Bar-Ilan University, Israel

[‡] Fondazione Bruno Kessler (FBK-irst), Italy

Taxonomy-based representations are widely used to model compactly large amounts of textual data. While current methods allow organizing knowledge at the lexical level (keywords/concepts/topics), there is an increasing demand to move towards more informative representations, which express properties of concepts and relations among them. This demand triggered our research on statement entailment graphs. In these graphs, nodes are natural language statements (propositions), comprising of predicates with their arguments and modifiers, while edges represent entailment relations between nodes. In this talk we report initial research that defines the properties of entailment graphs and their potential applications. Particularly, we show how entailment graphs can be profitably used for both knowledge acquisition and text exploration.

Beyond providing a rich and informative representation, statement entailment graphs allow integrating multiple semantic inferences. So far, textual inference research focused on single, mutually independent, entailment judgments. However, in many scenarios there are dependencies among Text/Hypothesis pairs, which need to be captured consistently. This calls for global optimization algorithms for inter-dependent entailment judgments, taking advantage of the overall entailment graph structure (e.g. ensuring entailment graph transitivity).

From the applied perspective, we are experimenting with entailment graphs in the context of the EXCITEMENT project industrial scenarios. We focus on the text analytics domain, and particularly on the analysis of customer interactions across multiple channels, including speech, email, chat and social media, and multiple languages (English, German, Italian). For example, we would like to recognize that the complaint they charge too much for sandwiches entails food is too expensive, and allow an analyst to compactly navigate through an entailment graph that consolidates the information structure of a large number of customer statements. Our eventual applied goal is to develop a new generation of inference-based text exploration applications, which will enable businesses to better analyze their diverse and often unpredicted client content. This task will be exemplified with data collected from real customer interactions, while referring to the EXCITEMENT Open Platform that we developed as a generic open source framework for textual inferences.