1. Publishable summary

EXploring Customer Interactions through Textual EntailMENT

Project funded by the European Community's Seventh Framework Programme
(FP7/2007-2013) under grant agreement n° 287923
http://www.excitement-project.eu/

EXCITEMENT is a 3-year research project (1/2012-12/2014) funded by the European Commission under FP7. The project consortium includes NICE Systems LTD (Israel) as coordinator, four academic partners, University of Bar Ilan (Israel), DFKI (Germany), FBK (Italy), University of Heidelberg (Germany), and two industrial partners, Almawave S.R.L. (Italy) and OMQ GmbH (Germany).

The main topic of the project is identifying semantic inferences between text units, a major language processing task, needed in practically all text understanding applications. While such inferences are broadly needed, there are currently no generic semantic engines or platforms for broad textual inference. The primary scientific motivation for the EXCITEMENT project is to change this ineffective state of affairs and to offer an encompassing open source platform for textual inference. On the industrial side, EXCITEMENT is focused on the text analytics market and follows the increasing demand for automatically analyzing customer interactions, which today cross multiple channels including speech, email, chat and social media.

There are two interleaved high-level goals for the project, which would yield two corresponding outcomes. The first is to set up, for the first time, a generic architecture and a comprehensive implementation for a multilingual textual inference platform and to make it available to the scientific and technological communities. To a large extent, the idea is to follow the successful experience of the Moses open source environment for machine translation, which has been making substantial impact on research in that field. This will enable developers of many text-processing applications to leverage the platform and boost their semantic inference capabilities. It will also provide developers of inference technology an effective environment for implementing and evaluating their components, and an easy entry-point for research in this field.
The second goal of the project is to develop a new generation of inference-based industrial text exploration applications for customer interactions, which will enable businesses to better analyze and make sense of their diverse and often unpredicted client content. These goals will be achieved for three languages, i.e. English, German and Italian, and for three customer interaction channels, i.e. speech (transcriptions), email and social media.

A major result of the project so far (i.e. month 24) is the first release of the EXCITEMENT Open Platform (EOP-1.0.1). We started with three entailment systems developed by the project partners: BIUTEE from Bar Ilan, TIE from DFKI, and EDITS from FBK, which have been migrated to the EOP architecture. The platform aims to automatically check for the presence of entailment relations among texts. It is based on a modular architecture and provides support for the development of algorithms that are language independent to a high degree. Thus, it allows developers and users to combine linguistic pipelines, entailment algorithms and linguistic resources within and across languages with as little effort as possible. As an example, a classification-based entailment algorithm can use, both separately and in combination, the results of a distance component and the

results of a BoW component, with the possibility to use lexical resources (e.g. wordnets) in different languages. The result is an ideal software environment for experimenting and testing innovative approaches for textual inferences.

EOP is distributed as open source software (http://hltfbk.github.io/Excitement-Open-Platform/) and its use is open both to users who are interested in integrating inference technology in applications, and to developers who are willing to extend the current functionalities.

Achievements in the first project year

During the first year of the project (January-December 2012) the EXCITEMENT consortium has progressed toward its goals, with significant achievements:

- The use cases of the project have been defined and two main scenarios have been identified: (i) text exploration based on entailment graphs; (ii) retrieval of customer interactions based on textual entailment.
- Collection of datasets of customer interactions, definition of the guidelines for the annotation of entailment graphs, realization of various annotated datasets for the three languages and the three channels of the project.
- Complete specification of the Excitement Open Platform architecture and implementation of the Java interfaces.
- Implementation of the Excitement Open Platform (EOP) specifications and initial population of the platform with linguistic modules, entailment decision algorithms and lexical resources. First EOP prototype.
- Collection of resources for the three languages of the project, and investigation of novel methodologies for the automatic acquisition of knowledge.
- Definition of the transduction layer for the integration of the EOP within the industrial scenarios provided by the industrial partners.
- Initial setting for the open source distribution of the platform, as well as publication and dissemination activities.

Achievements in the second project year

During the second year of the project (January-December 2013) the EXCITEMENT consortium has progressed toward its goals, with significant achievements:

- A collection of datasets of customer interactions in the three languages, needed for both development and testing purposes, has been fully annotated according to the entailment graph specifications defined in the first year.
- Updated specification of the Excitement Open Platform architecture and implementation of corresponding Java interfaces.
- Successful conclusion of the migration of the three initial entailment systems into the EOP platform. First public version (EOP 1.0.2 – August 2013) of the platform.
- Innovative research in textual inferences, with corresponding top level publications, focusing several aspects of the component-based approach proposed by EXCITEMENT.
- Collection of lexical resources for the three languages of the project, and investigation of novel methodologies for the automatic acquisition of knowledge.

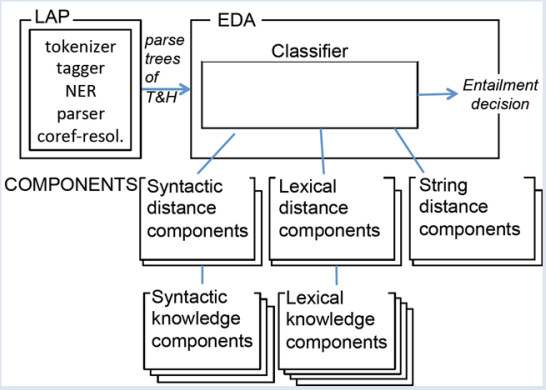
- First version of the transduction layer based on first year specifications.
- Evaluation of the EOP entailment algorithms based on a dataset (in three languages) of entailment text pairs (i.e. each containing a premise and a conclusion).
- First open source release of the EOP, including installation procedures, documentation, mailing list, distribution policies. Initial beta-testers and contributors.
- Initial design for the integration of the Transduction layer into the actual software environment of the project companies.
- Successful organization of the first EXCITEMENT workshop in Trento (November 2013), shaped as the Joint Symposium on Semantic Processing, with the participation of about 15 invited speakers and more than 80 researchers.
- Several relevant dissemination activities related to community building including a tutorial at AAAI and a full course at a Fall School.

Diagrams or photographs illustrating and promoting the work of the project

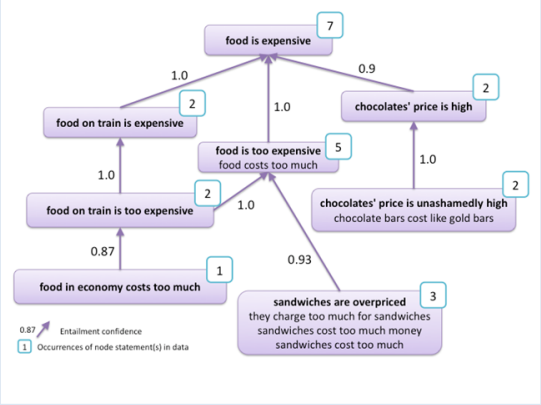
From the Excitement brochure published by the Excitement team members during META-FORUM2013.

EXCITEMENT Open Platform

- A modular architecture for textual entailment inferences
- Provides support for the development of language independent algorithms
- Easy combination of linguistic pipelines, entailment algorithms and linguistic resources within and across languages
- Ideal for experimenting textual inference methods, both for researchers and users
- Open source distribution, including state-of-the-art tools and resources for semantic inferences



http://hltfbk.github.io/Excitement-Open-Platform/



Entailment Graphs

- A powerful technology for representing complex relations in data streams as a hierarchy of statements
- Statement-level analysis: going beyond keyword-based techniques
- Graphs built from text with the EXCITEMENT Open Platform technology
- Unsupervised approach for content analytics, i.e. no need to train on specific (and limited) set of relations
- Two use cases: exploration of customer interactions and semantic retrieval of customer interactions

Academic partners:
· Bar Ilan University, Ramat Gan, Israel (I. Dagan)
· DFKI, Saarbrücken, Germany (G. Neumann)
· Fondazione Bruno Kessler, Trento, Italy (B. Magnini)
· University of Heidelberg, Germany (S. Pado)

Industrial partners:
· NICE Systems LTD, Ra'anana, Israel (project coordinator)
· OMQ GmbH, Berlin, Germany
· AlmaWave S.R.L., Rome, Italy

Project start date: 1/1/2012
Duration: 36 months