CROMER

a Tool for CROss-document Main Event and Entity Recognition

Christian Girardi, Manuela Speranza, Rachele Sprugnoli, Sara Tonelli

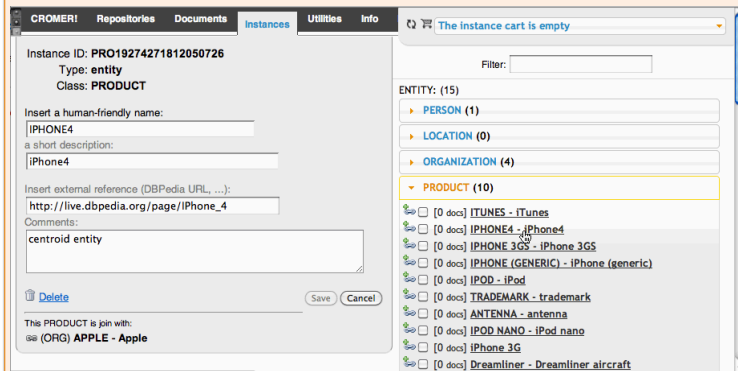cgirardi | manspera | sprugnoli | satonelli @fbk.eu

- objective -

CROMER is a web-based tool to manually annotate event and entity coreference across clusters of documents. It is designed to support annotators in linking entities and events to DBPedia and Wikipedia, so as to facilitate the automatic retrieval of additional semantic information. The tipical annotation procedure is the following:

- instance manager -

1. Before starting the annotation it is possible to customize an external semantic representation of instance types (e.g. Event, Entity, ...). Instances can also be subcategorized in classes (for instance, we can have entities of class Person, Organization, ...).
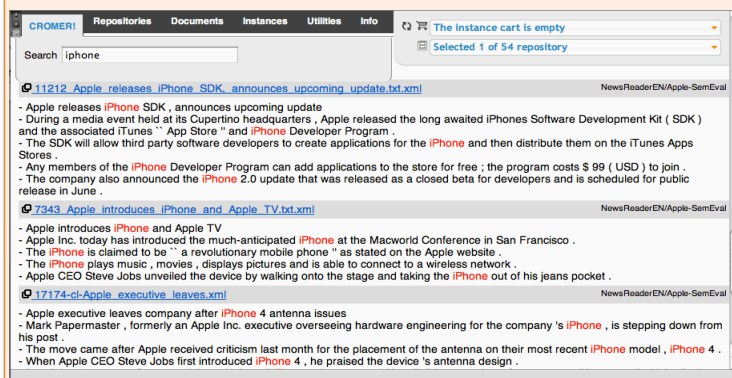

- features -

- create new types and classes for entities and events
- fill in entity and event templates
- link to external knowledge base sources
- establish relations among entities and events
- upload an existing set of instances from a CSV file
- share the information on the entities among all annotators

- trigger finding -

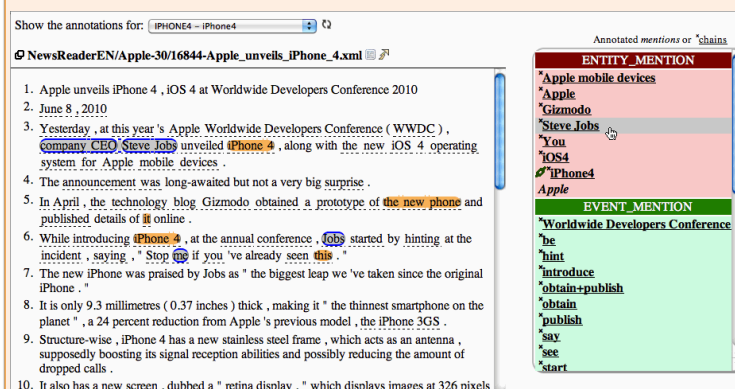**2.** The annotator searches the collection of documents containing the mentions of interest.

- features -

- upload the relevant docs distributing them into separated repositories
- support the search using single words, AND/OR operators, strings of tokens and wildchars
- export document in XML format
- avoid showing the already annotated documents
- get all the documents joined to an instance

- coreference annotation -

3. Each mention in the text can be linked to an instance producing intra and cross-document annotation.

- features -

- fast annotation using highlighting features
- annotation of discontinuous tokens is allowed
- two annotation modes are possible: one collaborative and one blind (without seeing the annotations done by other annotators)
- multiple assignments of the same token to different instances

- technical details -

- Tomcat/Java based web application
- Data stored in MySQL database and Lucene indexes
- Client realized using JQuery Javascript library
- Supports UTF-8 encoding

- use case -

In the NewsReader project, we observed that the average time needed to perform the annotation of a coreference is around 20 seconds, averaged over 4,000 assignments

CROMER is available on http://github.com/hltfbk/CROMER

Ready to use!