OntoNotes DB Tool

This is the first time in CoNLL that the data has many layers of annotation and there is significant supporting metadata, that we it is available in multiple format. One being the one that OntoNotes has realeased which is a separate file for each document and layer combination, in a hieararchical fashion. Second being the one mimicking the traditional CoNLL-style column format. We don't have any tools for researchers to manipulate the CoNLL-style format in, but for the standard OntoNotes format we have developed a Python API that makes it easier to read and manipulate it, and which we are making available as part of the CoNLL support software.

Documentation

The design of this API was discussed in the following article:

- **OntoNotes: A Unified Relational Semantic Representation**

   Sameer Pradhan, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw, Ralph Weischedel, invited paper, *International Journal of Semantic Computing*, Vol. 1, No. 4, pp. 405-419, 2007

And, there are several examples on using it, in the following tutorial presented at HLT/NAACL in 2009:

- **OntoNotes: The 90% Solution**

   Sameer Pradhan and Nianwen Xue *Presented at Human Language Technology Conference*, Boulder, CO, 2009.

The Python API can be downloaded below, and you can also download, or browse the documentation of the API [

][HTML]

This is still a work in process and we would welcome your feedback or comments at sameer@ cemantix.org

[OntoNotes DB Tool v0.999b r6778 (Beta)](#)   **Scorer**

We will be following the scoring strateey used in the ["SemEval Task: Coreference Resolution in Multiple Languages"](#)   .
There is an ongoing debate in the community on the way the performance of a corference system is evaluated. Originally, the MUC metric was the standard for several years. Since then three metrics have been proposed — B-CUBED, CEAF, and most recently BLANC. We will be scoring the output of the CoNLL coreference systems using all the metrics. One of them (we haven't finalized yet) will be used to determine the winner.

- Scorer:
- v1
- v2
- v3
- v4 (updated May 11, 2011)

This was the version used for the official evaluation. It has since been superseded by following versions.

- v7 (updated Dec 28, 2013)

This version fixed the computation of B-CUBED and CEAF metrics. Refer to [Pradhan et al. (2014)](#)   for details.

- v8 (updated July 16, 2014)

This version updated the BLANC scorer to handle predicted mentions. Refer to [Pradhan et al. (2014)](#)   and [Xi aoqiang et al. (2014)](#)
for details.

- [v8.01](#) (updated August 2, 2014)

This is a bugfix version that fixes a bug which made BLANC crash when the input contained the mention in multiple clusters. The reference implementation is available on [google code](#) . The participating systems from both years have been re-scored and the a
[spreadsheet](#)
has been made available. We will soon add these updated numbers to the webpage under a separate tab. This development has mostly affected the magnitude of the scores, and the overall rankings have been slightly affected. This development does not change the original top system rankings.