TwitIE (http://gate.ac.uk/wiki/twitie.html)

TwitIE is an information extraction tool that was built and developed using the full GATE architecture. TwitIE is an IE tool that is specifically made to work with information extraction from one of the noisiest dataset in the world, that is, social media data specifically, Twitter data. TwitIE makes use of the existing modules of the GATE architecture together with some modified ones to enable it to perform the following features/functionalities:

* Custom Language Identifier to support language identification for social media/Twitter data;
* Twitter Tokenizer to enable the system to properly handle smilies, usernames, hashtags, URL’s and etc.;
* Twitter POS Tagger (also available as a standalone tool) to enable the system to properly perform POS tagging in Twitter data; lastly,
* Text Normalization PR to enable the system to perform correction of slangs and misspellings in the tweet.

TwitIE is available as a standalone application from the GATE Twitter plugin. TwitIE makes use of the Twitter and Stanford Tagger plugins as its baseline and reference for performing the information extraction process.