

Automatic Translation in Social Internetworking

By Michael J. Watts (<http://mike.watts.net.nz>) created 31 August 2011, last modified 13 September 2011.

This report briefly describes how the multi-lingual social media presences of the IEEE Computational Intelligence Society were set up. The goal of this work was to automatically translate and propagate posts on the CIS blog:

<http://ieee-cis.blogspot.com/>

There are two actions associated with this. The first is the translation of the blog itself. This was achieved simply by enabling the Google Translation gadget, which uses the Google translation engine to translate the entire blog. The language must be selected by the blog visitor, although they will be prompted to do this if the default language of their browser is set to anything other than English.

The second action is translating the posts that are then automatically propagated to other social media platforms. For English, this is done as per the description in my previous report on social internetworking, available here:

<http://mike.watts.net.nz/SocialInternetworking.pdf>

This method cannot be used as-is for this, however, because the RSS feed from the blog, is only available in English. That is, while the blog itself can be viewed in other languages, the RSS feed cannot. The solution to this problem is to access the blog RSS feed via Yahoo! Pipes. This service allows for various manipulations of an RSS feed, including translating the items in the feed. A Yahoo! Pipe will accept a RSS feed from a nominated source, process it, and emit an RSS feed. Unfortunately, Yahoo! Pipes has fewer translation options than Google Translate. The twelve available languages are:

- Chinese (Simplified)
- Chinese (Traditional)
- French
- Portuguese
- Dutch
- German
- Spanish
- Greek
- Italian
- Russian
- Korean
- Japanese

This is compared to the more than fifty languages available from Google Translate. A separate pipe was constructed for each of these languages (with the exception of Traditional Chinese). The general structure of each pipe is illustrated by the example of the Simplified Chinese pipe, shown in Figure 1. The pipes for the other languages are similar, with only the languages to which they are translating being different. Since the blog posts are in English, only English is specified as the source language in the translation modules.

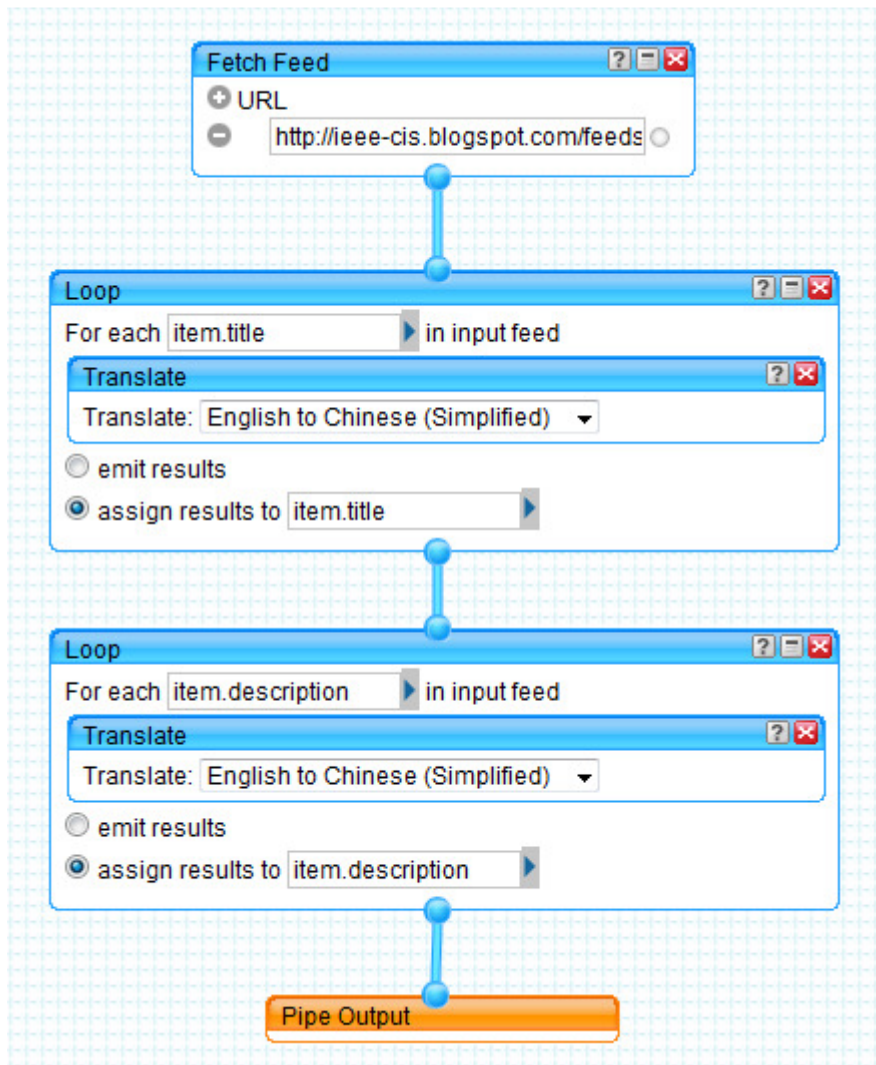


Figure 1 - pipe to translate blog RSS feed into Simplified Chinese

Once posts have been translated into the target language by the relevant pipe, the resulting RSS feed can be used to feed the posts to the connected social media sites. Although most of these sites allow posting in multiple languages and character sets, it would be over-whelming for readers if the same post appeared in multiple languages on the same site. Therefore, new presences were established on each site for each language. This approach requires more setup work, but presents a much cleaner interface for users, and makes it easier to integrate language-specific sites into the process.

The general way in which information flows from the blog to the final sites is shown in Figure 2. There are no language-specific sites shown in this figure.



Figure 2 – Data flow for any language

There are several Chinese language sites in operation: Weibo, Zuosa and Digu. As Weibo cannot directly connect to an RSS feed or an aggregator site like hellotxt.com, Weibo must get its posts from another site like Tumblr. The way in which data flows from the blog for Simplified Chinese is presented in Figure 3, where the Chinese-specific sites are also shown.



Figure 3 – Data flow for Chinese translations

So far, Chinese, French, Portuguese, Korean, Greek, German and Spanish sites have been set up. The current sites are listed by language below, but not all of them are currently receiving updates!

Chinese (simplified)

- <http://twitter.com/#!/ieeeciscn>
- <http://ieecis-cn.tumblr.com>
- <http://shoutitout.shoutem.com/ieeeciscn>

<http://ieeeciscn.jaiku.com>
<http://www.plerb.com/ieeeciscn>
<http://www.qaiku.com/home/ieeeciscn>
<http://www.plurk.com/ieeeciscn>
<http://identi.ca/ieeeciscn>
<http://zuosa.com/ieeecis>
<http://digu.com/opi37q>
<http://weibo.com/ieeecis>

French

<http://twitter.com/#!/ieecisfr>
<http://ieecisfr.tumblr.com>
<http://shoutitout.shoutem.com/ieecisfr>
<http://ieecisfr.jaiku.com>
<http://www.plerb.com/ieecisfr>
<http://www.qaiku.com/home/ieecisfr>
<http://www.plurk.com/ieecisfr>
<http://identi.ca/ieecisfr>

Portuguese

<http://twitter.com/#!/ieecispt>
<http://ieecispt.tumblr.com>
<http://shoutitout.shoutem.com/ieecispt>
<http://ieecispt.jaiku.com>
<http://www.plerb.com/ieecispt>
<http://www.qaiku.com/home/ieecispt>
<http://www.plurk.com/ieecispt>
<http://identi.ca/ieecispt>

Korean

<http://twitter.com/ieeciskr>
<http://ieeciskr.tumblr.com>
<http://shoutitout.shoutem.com/ieeciskr>
<http://ieeciskr.jaiku.com>
<http://www.plerb.com/ieeciskr>
<http://www.qaiku.com/home/ieeciskr/>
<http://www.plurk.com/ieeciskr>

Greek

<http://twitter.com/ieecisgr>
<http://ieecisgr.jaiku.com>
<http://www.plerb.com/ieecisgr>

<http://ieecisgr.tumblr.com>
<http://shoutitout.shoutem.com/ieecisgr>
<http://www.qaiku.com/home/ieecisgr/>

German

<http://twitter.com/ieecisde>
<http://ieecisde.jaiku.com>
<http://www.plerb.com/ieecisde>
<http://ieecisde.tumblr.com>
<http://shoutitout.shoutem.com/ieecisde>
<http://www.qaiku.com/home/ieecisde/>

Spanish

<http://twitter.com/ieecises>
<http://ieecises.jaiku.com>
<http://www.plerb.com/ieecises>
<http://ieecises.tumblr.com>
<http://shoutitout.shoutem.com/ieecises>
<http://www.qaiku.com/home/ieecises/>