

Nanyang Technological University Update

Francis **Bond**
and many, many more

Computational Linguistics Lab (凡士研)
Linguistics and Multilingual Studies,
Nanyang Technological University

[<bond@ieee.org>](mailto:bond@ieee.org)

DELPH-IN 2018-06-18



- NTU

- ▶ Wordnets: Multi, Jpn, English, Chinese, Bahasa, *Myanmar, Kristang*
 - ★ Collaborative Interlingual Index
 - ★ hosted GWC 2018
- ▶ Grammars:
 - ★ Indonesian (INDRA)
 - ★ Chinese (Zhong [||]: Mandarin and Cantonese)
 - ★ Japanese (JACY) — bug fixes, minor enhancements
- ▶ Tools:
 - ★ OMW version 2.0, LMF checking
 - ★ intsem.fx, visualkopasu (Integrated Semantic Framework)
- ▶ Intelligent Tutoring and Feedback on Errors
- ▶ English Essay writing help

Open Multi-Lingual Wordnet

- Collaborative InterLingual Index (CILI)
 - ▶ Initial testing pretty much done
 - ▶ Wow, there are a lot of interesting things you can do
 - ★ And we check for/document many of them
- More languages and words
 - ▶ Sesotho, Tswana, Venda, Xhosa, Zulu, Gaelic, Abui, Moroccan Arabic, *Turkish, Czech*, ...(40+ curated languages)
 - ▶ National Cancer Institute Thesaurus, Colloquial Wordnet
 - ▶ 10,000+ new English synsets (enwordnet)
- More links
 - ▶ Interlanguage sense level links (Harmony Project)
 - ▶ Better sentiment

NTU multilingual corpus

- Small, deeply analysed corpus
 - ▶ 6,000+ sentences x 3 languages (cmn, eng, jpn); 2,000 in Indonesian
 - ★ Mainichi Newspaper (NICT translations)
 - ★ Sherlock Holmes (**DANC**, **SPEC**; **REDH**, **SCAN** [eng])
used in ACL workshop on Dialogue
 - ★ Cathedral and the Bazaar (many languages)
 - ★ Spider's thread 「蜘蛛の糸」
 - ★ Singapore Tourist data (plus Korean, Viet, Indo)
 - ▶ **Hand alignment**, **WordNet tagging**, Treebanking, Sentiment, ...
- Revised the DB structure; second round of wordnet tagging
 - ▶ Much behind schedule
 - ▶ Wordnet extension is being done in parallel
- Tagged Italian (**SPEC**)
ready to start Polish, Bulgarian, Dutch, German, Abui (shared task)

Newish Projects

- A collaborative network for usage-based research on lesser-studied languages (JSPS: Tokyo University of Foreign Studies) — building large Indonesian corpus
- Digital Mapping the Literary Epigraph: Quantitative analysis of literary influence using network theory and thousands of epigraphs (NTU)
- *Digital Humanities Cluster* at the College of Humanities and Social Sciences
- Year four of research with Fuji-Xerox
 - ▶ WSD, Sentiment and transfer learning

Members

- Staff:
E umamaheswari Vasanthakumar, Takayuki Kuribayashi, Wenjie,
Irfan, ...
- Visitors/Interns:
Hiroki Nomoto
- Students:
Zhenzhen, David, Tuan Anh, Mike, Jim, Luis, Hannah, Noblas, Liu
- Graduated/ing
 - ▶ David Moeljadi, 2018 (PhD: LMS, NTU) *INDRA JATI: an Indonesian Resource Grammar and a Treebank for Indonesian*
 - ▶ Elvis Albertus bin Toni, 2018 (PhD: LMS, NTU) *Emotions in Adonara-Lamaholot*
 - ▶ Michel Goodman, 2018 (PhD: Linguistics, University of Washington, USA) *Semantic Operations for Transfer-based Machine Translation*
 - ▶ Jim Breen, 2018 (PhD: Engineering, University of Melbourne, Australia) *Extraction of Neologisms from Japanese Corpora*

