

Summary

- Parallel data are vital for MT.
The more and better, the better.
- Several projects for document alignment.
Project suggestion: Targeted re-crawl based on Kúdela et al. (2017).
- Sentence alignment “solved”.
- Word alignment ill-defined but used to be very important.
Plus all the funny heuristics...
- Beyond word alignment:
 - Phrase alignment never got wide-spread; too tied to PBMT anyway.
 - T-Alignment costly (T-layer needed).
 - Project suggestion: NN LEAF.