

データベース名称： 日中特許用語辞書

I データの概要

日中特許用語辞書を、既存の対訳辞書を使用せず、日中特許対訳コーパスをもとに日本語漢字から中国語簡体字への文字マッピング知識や既存の統計的機械翻訳、用語抽出等の各種自然言語処理ツールを用いて自動構築し、最後に人手による修正作業を行って整備した。

II 関連発表

- Keiji Yasuda, Eiichiro Sumita. Building a Bilingual Dictionary from a Japanese-Chinese Patent Corpus. In Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Karlovasi, Samos, Greece, March 24-30, 2013, Proceedings, Part II, pages 276-284, March.
- 安田圭志, 隅田英一郎. 日中特許対訳コーパスを用いた対訳辞書の自動構築. 言語処理学会第 19 回年次大会発表論文集, pp. 306-309, March 2013.

III データ内容

- 日中特許用語対訳辞書
日本語の見出し語 169959 語、中国語の見出し語 169384 語からなる。各日本語に対し 1 ないし複数の中国語が対応し、フォーマットは以下になる。

日本語<tab>中国語 1 <tab>中国語 2 <tab>…<tab>中国語 n

以下に辞書の例を挙げる。

電子クーポン	e 优惠券	优惠券			
電子コネクタ	电气连接器	电子连接器			
電子コントローラ	电子控制器	电子控制			
電子コンポーネント	电子部件	电子组件	电子元件	电子元器件	
電子サイクロトロン共鳴	电子回旋共振				
電子システム	电子系统				