

採択事業概要

対象事業名	機械学習を活用した非アクセシブルな PDF 文書の構造化とテキスト抽出に関する研究開発
対象事業者	イースト株式会社
<p>PDF 文書の版面の構成要素を機械学習によって判定し、見出しの識別や柱・ノンブル等の不要な情報の除去を行って構造化する。構成要素に含まれるテキストは OCR と PDF 解析プログラムを併用して高い精度で抽出し、障害者や高齢者等の持つさまざまなニーズに対応できる形式で提供する。</p>	