平成27年度 情報工学コース卒業研究報告要旨

大西 研究室	氏　　名	荒 井　秀 育
卒業研究題目	OCRを用いた英文中の略語の意味推定システム	

背景と目的

書籍や論文等の紙面に印刷された英文を読んでいるとき，略語が使用されていることが多々ある．略語の意味は初出の際に説明されることが多いが，未説明のこともある．また，読み進み，定義されたページを探すことも考えられる．このように，意味解説のない未知の略語を目にする機会は少なくない．略語の意味を調べる際，同じ略語が複数の意味を持つことも多い．さらに，略語が代表的な意味を持つ場合は，それ以外の意味についての情報は得づらい．このような点から，略語の意味の特定を支援するシステムが望まれる．そこで，本研究ではOCRを用いて，英文中の略語の意味推定を行うシステムを提案する．

提案システムの概要

本システムはAndroidアプリケーションとして実現した．OCRを用いることにより，略語と同時にその周辺の文章を取得し，意味の推定に利用する．アプリケーションを起動し，カメラ機能を利用して文章を撮影する．その画像にOCR処理を施すことで，画像中の文章のテキストデータが得られる．テキストデータから対象とする略語を選択し，意味推定を開始する．

意味推定処理では，Wikipediaの曖昧さ回避ページのHTMLファイルの構成を利用する．選択した略語をURLに組み込むことで，その略語の曖昧さ回避ページのHTMLファイルから，リスト構造タグに含まれる選択略語の意味候補を抽出する．次に，意味候補の定義文中の単語と，OCRで取得した文章中の単語の一致回数を算出し，これを評価値とする．推定結果として評価値の高い順に意味候補を並べ替えて表示する．必要に応じて，曖昧さ回避ページの内容を表示する．略語の意味が一意である場合には，取得したHTMLファイルの内容を表示する．

実験と結果

意味推定処理の精度を調査する実験を行った．この実験では“IMF”を対象の略語とし，“IMF”を含む論文100編のテキストデータを入力とした．評価値が最大の意味候補と正しい意味が一致したのは60編だった．

次に，撮影処理とOCR処理が意味推定システムに与える影響を調査する実験を行った．テキストデータを直接与えた推定結果と，画像を撮影しOCR処理を施して得たテキストデータを利用した推定結果を比較した．前実験の1位の論文60編を用いて比較を行った結果，推定結果は55編で一致した．

表1：意味推定精度実験の結果

1位（単独）	1位（複数）	2位以下	順位外	該当無し
53	7	3	17	20

表2：提案システム精度実験の結果

1位（単独）	1位（複数）	2位以下	順位外	該当無し
50	8	0	1	1