

## 平成27年度 情報工学コース卒業研究報告要旨

山本 研究室	氏 名	若 林 丈 紘
卒業研究題目	要求仕様書と設計書における語彙の差分と出現頻度の分析	
<p>ウォーターフォール型のソフトウェア開発では、要求仕様書、基本設計書といったドキュメントが作成される。開発者は利用者の要求に基づいてこれらのドキュメントを作成し、その後ドキュメントに従ってコーディングやテストを行う。開発者は、基本設計書がその前段階で記述された要求事項を漏れなく満たしているかどうか、要求仕様書の記述を実現する部分に対応付ける必要がある。多くのドキュメントは自然言語で記述されるため、これらは目視で評価することが一般的である。目視による評価は作業コストが大きいという問題がある。このことから、自然言語処理技術によって要求文と設計文の対応付けを支援する手法が提案されている。既存の手法では、要求文と設計文の類似度を算出し、ある閾値を超えたかどうかで対応関係の有無を判定する。閾値は分析者の経験や感覚に基づき決められる。このため、検出できる記述漏れの内容や割合は分析者の能力に依存する。</p> <p>本研究では、分析者に依存しない方法による、設計漏れ要求に含まれる語彙の自動抽出を目指す。設計漏れ要求は、対応する設計文が基本設計書に存在しない要求文を指す。抽出された語彙をドキュメントのレビューにあらかじめ提示することで、設計漏れ要求を検出する作業コストを減らすことができる。本稿では、設計漏れ要求には要求仕様書と基本設計書の語彙の差分が含まれる、と仮説を立て、4組の開発ドキュメントを対象に3つの調査を行った。</p> <p>まず、実際に語彙の差分をもとにドキュメント中にある設計漏れ要求を検出できるかを確かめる。4組すべてのドキュメントにおいて、要求仕様書には存在するが基本設計書には存在しない語彙があることを確認した。これらの語彙のうち、出現頻度の大きい語彙を含む要求文を目視により調べた結果、4組すべてのドキュメントにおいて、設計漏れ要求が見つかった。4組合わせて計303語の差分の語彙に対し、基本設計書に対応する記述がない要求文が60件見つかった。ドキュメントの差分語彙をもとに設計漏れ要求を検出できた。</p> <p>差分語彙をもとに記述漏れを検出する上では、検出できる設計文と検出できない設計文の内容を、あらかじめ把握する必要がある。基本設計書から章や節を欠落させることで疑似的に記述漏れのある基本設計書を用意し、どのような設計文であれば記述漏れを語彙の差分によって検出できるかを調べた。調査の結果、4組すべてのドキュメントにおいて、設計漏れを検出できる語彙が基本設計書の一部に偏在していることが分かった。開発者が語彙の差分をもとに設計漏れ要求を探す場合、漏れを検出できない語彙をあらかじめ把握することで、他の部分を検出する作業コストを削減できると考えられる。</p> <p>ある章節の情報を知る上で重要な手がかりとなる単語は、ドキュメントにおいて局所的に偏在するため、その章節の記述が漏れたときに差分語彙となりやすい。重要語が多く存在する章節の内容が分かれば、差分語彙で検出できる記述漏れの内容を推測することができる。そこで、基本設計書の各章節に存在する単語のTF-IDFを算出し、各章節における重要語の分布を調べた。調査の結果、4組すべてのドキュメントにおいて、システムの構成要素やシステム開発に固有の用語が頻出する章節では、重要語が多く存在することが分かった。一方、システムのデータ項目や業務ドメインに固有の用語が頻出する章節では、重要語がほとんど存在しないことが分かった。基本設計書の各章節における単語のTF-IDFは、差分語彙によって記述漏れを検出できる章節であるかを判断するための指標として利用できると考えられる。</p>		