

論文内容の要旨

博士論文題目 Unlocking Software Documentation: Sentiment Classification and On-hold Self-Admitted Technical Debt Identification

氏 名 MAIPRADIT RUNGROJ

Software documents refer to all written documents in software development, which play a crucial role in knowledge sharing between software developers. Despite the benefits of software documentation, document creation and maintenance are frequently overlooked. Many software documents are often outdated, and many small to medium software projects have little to no software documentation.

This thesis presumes that the problem of little to no software documentation and outdated documents can be tackled by unlocking software documents to show the benefit of existing information. To address this, this thesis first unlocks software document information by proposing a new technique for accessing existing information on software engineering data sets using sentiment classification. The results show that using automated machine learning with n-gram inverse document frequency shows promising results in tackling this problem. Second, to reduce outdated documents by finding what kinds of tasks are amenable to automated management. And found one particular class of debt amenable to automated management: on-hold SATD, i.e., debt which contains a condition to indicate that a developer is waiting for a certain event or an updated functionality to be implemented elsewhere. Third, this thesis investigates the potential of removing outdated documents by automatically detecting on-hold SATD and identifying its condition. The results show that the proposed design can reliably identify on-hold SATD and also mine the issue tracker to check if the On-hold SATD instances are “superfluous” and can be removed.

In all, this thesis emphasizes the benefit of unlocking software documents for software development. Furthermore, this thesis provides practical implications for extracting information from software documents through sentiment classification and for managing outdated software documents through suggesting the removal of outdated on-hold SATD comments.

(論文審査結果の要旨)

本論文は、ソフトウェア開発における文書作成（ドキュメンテーション）を2つの視点で解析することで、ソフトウェアの設計や実装に関する情報や知識を開発者間で共有するという、文書作成本来の目的をより効果的、実践的に実現する手法を示すものである。ソフトウェア開発では、ソフトウェアの設計や実装に関する情報や知識を示す多様な文書（ソフトウェア文書）が作成される。ソフトウェア文書による情報・知識共有の有用性は広く認められているが、その一方で、明示されている情報・知識が不足していたり、ソフトウェアの実態から乖離していたりする場合も多いことが指摘されている。

本論文では、ソフトウェア開発を行う上でより有益な情報をソフトウェア文書に付加するとともに、更新されずソフトウェアの実態をもはや表していない情報を特定、削除することで、ソフトウェア文書による情報・知識の共有の実効性を高めようとするものである。具体的には、感性分析（Sentiment Analysis）を用いることで、ソフトウェア文書に含まれる文章の肯否を分類し、当該文書の理解や解釈を容易にする手法を開発した。更に、ソースコード中の自己申告型技術的負債（Self-admitted Technical Debt: SATD）のサブクラスとして「保留中の自己申告型技術的負債（On-hold SATD）」を新たに定義し、負債ではなくなったSATDを特定、削除する手法を開発した。開発手法において対象とするソフトウェア文書は、「設計書」、「ソースコード中の注釈（コメント）」、および、「開発者が意見交換を行うフォーラムでやり取りされたメッセージ文」である。オープンソースソフトウェアにおいて作成されたソフトウェア文書に開発手法を適用した結果、その有効性、妥当性が確認された。

以上のとおり、本論文は、ソフトウェア開発で広く用いられているソフトウェア文書をこれまでにない視点で捉えた研究であり、当該文書によるソフトウェア開発者間での情報・知識共有の実効性を高め、その活用を促進する。ソフトウェア文書に有益な情報を付加する一方で不要な情報を特定、削除する手法、そして、それらから得られた知見は、広くソフトウェア開発とソフトウェアエコシステムの高度化、そして、ソフトウェア工学研究の発展に大きく貢献することから、博士（工学）論文として価値あるものと認める。