

自然言語処理技術を用いた 上流工程の品質向上アシストシステム構築に向けて

キヤノン株式会社

秦 理貴

hata.yoshitaka@mail.canon

設計資産の活用の問題点

設計資産(設計書と障害票)を設計知として体系的に蓄積し利用することが上流工程の品質向上のアシストにつながる

＜問題点＞

- ・設計知の構築ために経験則が必要で属人化
- ・自然言語で記述された膨大な設計資産から関係資産の検索の困難性

自然言語処理による設計知構築

設計書と障害票に対して自然言語処理技術の適用による人手に頼ることのない設計知構築の提案

設計知: 設計書と障害票の関連度を数値化

- ・TF-IDFによる特徴ワード抽出
- ・特徴ワードを用いたコサイン類似度算出

提案手法の概要

関連1: 特徴ワードの抽出

設計書と障害票の特徴付けるワード抽出をおこない、障害票を検索できる準備をおこなう

関連2: 設計書と障害票の関連度の抽出

設計書と障害票の共通特徴ワードを用いてコサイン類似度による関連度の抽出をおこなう

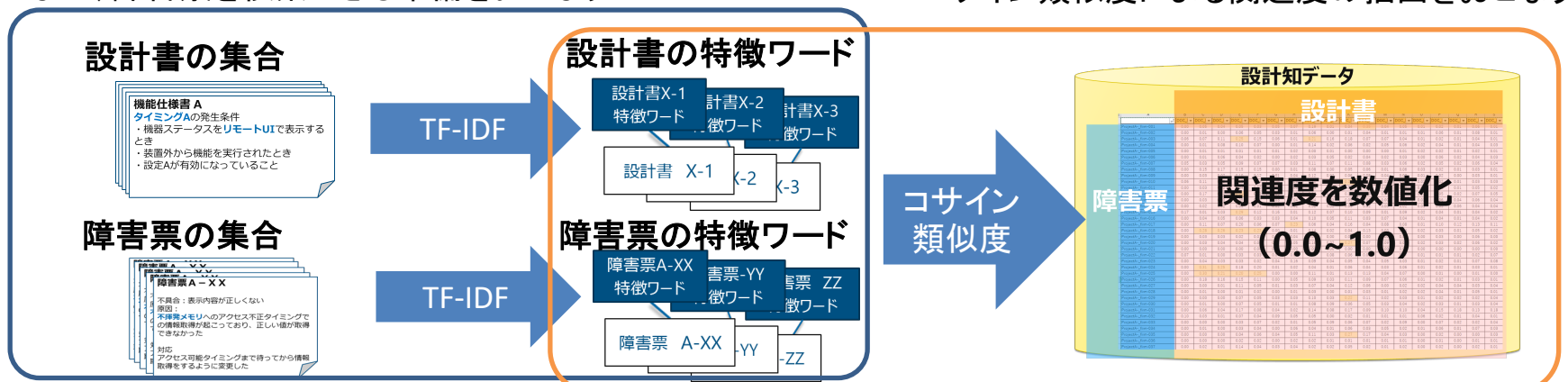


図1. 設計書と障害票からの設計知データの構築方法

評価

検証1: 特徴ワード抽出の妥当性検証(関連1)

提案手法で抽出した特徴ワードと人手で抽出した特徴ワードを比較し約80%が一致

提案手法で特徴ワード抽出が可能

検証2: 設計書と障害票の関連度の妥当性検証(関連2)

算出されたコサイン類似度が高いものほど関連づけられた障害票が妥当であることが確認できた

活用できる設計知データ構築が可能

	A: 障害タイトル	B: 現象	C: 原因	D: 対策
期待するワード	W-A002 W-A004 W-A100	W-A002 W-A004 W-A100	W-A004 W-C001	W-D100 W-C001 W-D101
抽出した特徴ワード	W-A002' 0.50 W-A002 0.50 W-A002' 0.44 W-A004 0.23 W-A005 0.21	W-A005 0.39 W-A002' 0.32 W-A002 0.32 W-B001 0.30 W-B002 0.25 W-A002' 0.22 W-A004 0.19	W-C001 0.52 W-C002 0.34 W-A004 0.30 W-A004' 0.24 W-A004' 0.24	W-D001 0.27 W-D002 0.24 W-C001 0.23 W-C002 0.21 W-D003 0.21

青字が一致している特徴ワード

図2. 特徴ワード抽出結果の一例

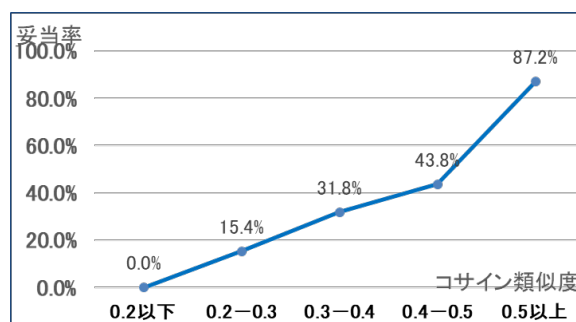


図3. コサイン類似度と妥当性の相関

今後の展望

1. 提案した設計知データを元に、新規設計書から障害票検索システムの構築
2. より複雑な関連抽出による、設計知データ構築

人が気づくことが難しい過去設計資産(障害票)へのアクセスを可能とし、品質向上につなげる