

ベクトル空間検索モデルへの 緊密用語集合の適用

Applying Family Terms to the Vector Space Retrieval Model

東京工芸大学工学部コンピュータ応用学科
宇田川 佳久

研究の最終目的(チャレンジ):

SW開発における文書管理を精確かつ
効率的にする

発表の流れ:

1. 研究の背景
2. 処理プロセスの全体像
3. SysMLによる要求・仕様記述
4. ベクトル空間検索モデルの改善
5. 実験結果と考察
6. 今後の研究方針

1. 研究の背景 (1/2)

- SW開発では、多種の文書が作成される
要件定義書、外部・内部設計書、ソースプログラム
- 各文書は独自のモデル(表現手段+抽象レベル)で記述されている
要件定義書 --- 自然言語、エンドユーザの視点
設計書 --- 図形言語(UML/SysML)、実装者の視点
プログラム --- プログラミング言語、機械の視点
- 表現手段や抽象レベルが違っていても同じものを表している
⇒ 各文書間の関連性(トレーサビリティ)を管理する手法を研究している

1. 研究の背景 (2/2)

本研究のアプローチ

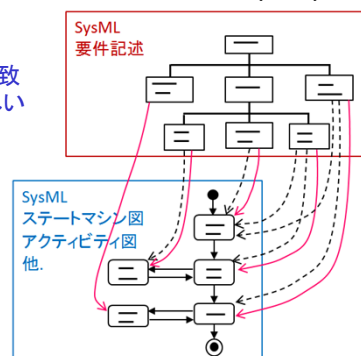
1. データマイニング手法を使った文書間のトレーサビリティの自動検索。
2. 文書は、業界標準である“SysML” に準じて記述する。
※ただし、SysMLに特化したアプローチではない

発表の流れ:

1. 研究の背景
2. 処理プロセスの全体像
3. SysMLによる要求・仕様記述
4. ベクトル空間検索モデルの改善
5. 実験結果と考察
6. 今後の研究方針

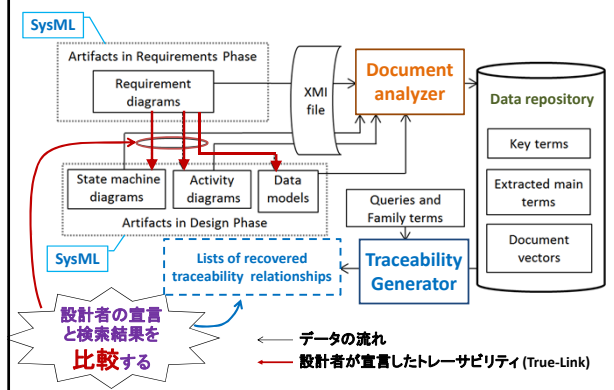
2. 処理プロセスの全体像(1/2)

実線と点線が一致
することが望ましい



実線: 設計者が宣言したトレーサビリティ
点線: 情報検索によって回復したトレーサビリティ

2. 処理プロセスの全体像(2/2)

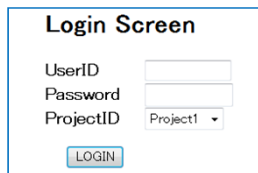


発表の流れ:

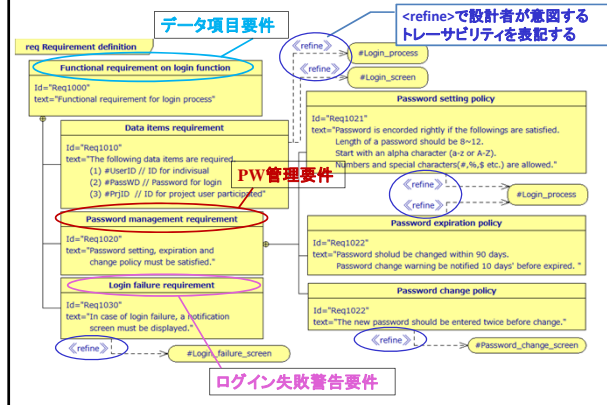
1. 研究の背景
2. 処理プロセスの全体像
3. SysMLによる要求・仕様記述
4. ベクトル空間検索モデルの改善
5. 実験結果と考察
6. 今後の研究方針

Webシステムへのログイン機能要件

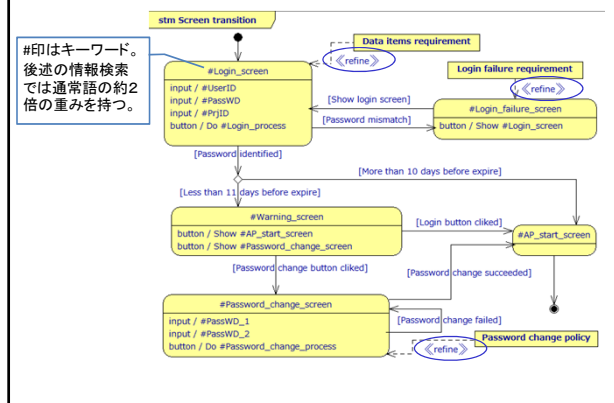
- 入力項目要件 ユーザID
パスワード
プロジェクトID
- パスワード管理要件(構成・変更ルール)
- ログイン失敗時の警告要件



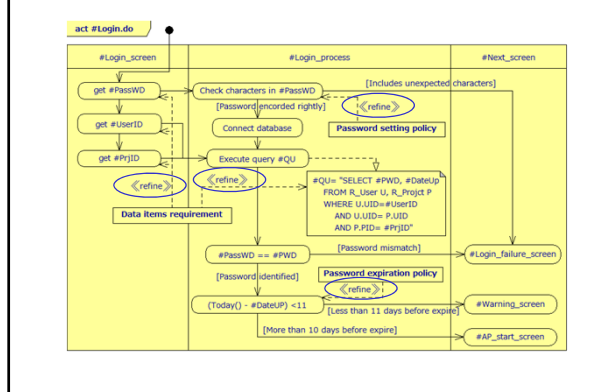
SysMLによる要求定義(階層化機能)



SysMLによる画面遷移設計(ステートマシン図)



SysMLによる処理フロー設計(アクティビティ図)



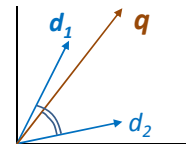
発表の流れ:

1. 研究の背景
2. 処理プロセスの全体像
3. SysMLによる要求・仕様記述
4. ベクトル空間検索モデルの改善
5. 実験結果と考察
6. 今後の研究方針

4. ベクトル空間モデル

$$d_j = (w_{1,j}, w_{2,j}, \dots, w_{N,j})$$

$$q = (w_{1,q}, w_{2,q}, \dots, w_{N,q})$$



$$\text{Similarity}(d_j, q) = \cos(d_j, q) = \frac{\sum_{i=1}^N w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^N w_{i,j}^2} * \sqrt{\sum_{i=1}^N w_{i,q}^2}}$$

ベクトル空間モデルの改良(1/2)**文書解析 (Document Analyzer)**

1. 抽出単語に、その単語が抽出された**文書のID**を付ける
 - 例:
 要求定義で抽出された単語には“Requirement”
 設計書で抽出された単語には“Design”
2. クエリを構成する主要語を“**緊密用語集合**”として扱う
 - 例:
 {A, B, C, D}がファミリー単語の場合、ドキュメント(検索対象)が{A, B, C}を含めば、{D}も含まれているものとして扱う(ある種の推論ルール)

ベクトル空間モデルの改良(2/2)**トレーサビリティ生成 (Traceability Generator)**

□ 下記のモードで動作する:

- (1) ベクトル空間検索モデル (VS)
- (2) 文書IDを加味したベクトル空間検索モデル (VA)
- (3) 緊密用語集合と文書IDを加味したベクトル空間検索モデル (VF+VA)

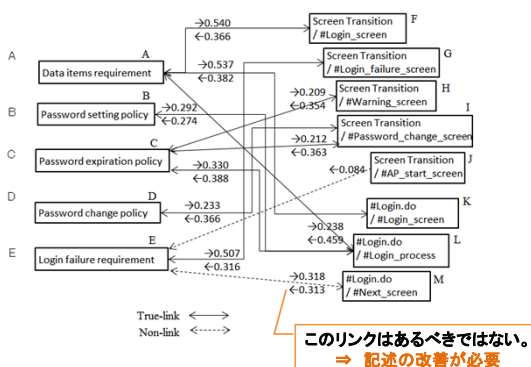
発表の流れ:

1. 研究の背景
2. 処理プロセスの全体像
3. SysMLによる要求・仕様記述
4. ベクトル空間検索モデルの改善
5. 実験結果と考察
6. 今後の研究方針

5. 実験結果と考察**実験の概要**

- 3個のSysML図を、13個の文書ベクトルに変換した
- 82個の用語を、文書ID付きで抽出した
- 文書ベクトルの生成には、**tf_idf**による重み付けを行った
 > **tf_idf**: term frequency_inverse document frequency
- 実験は双方向で行った
 > 要件から設計の検索
 > 設計から要件の検索

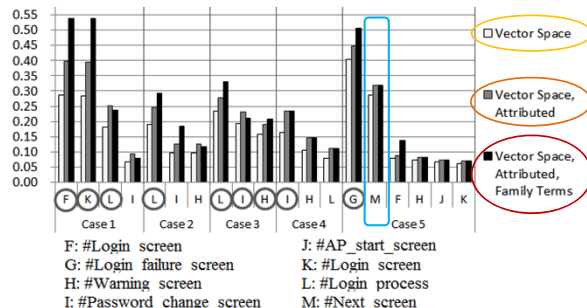
検索結果の概要



このリンクはあるべきではない。
⇒ 記述の改善が必要

5. 実験結果と考察

要件から設計の検索結果

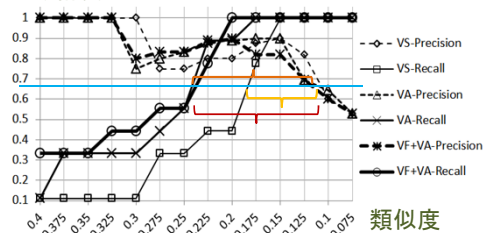


5. 実験結果と考察

再現率と精度とも0.67で類似度の範囲を評価した

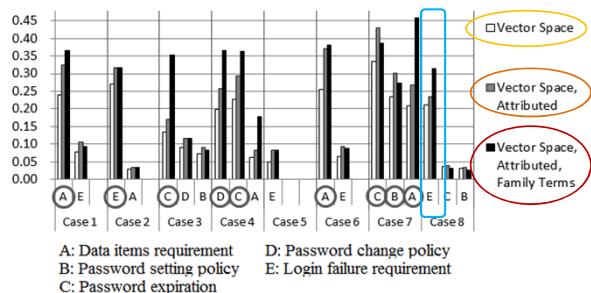
- VSモード: 類似度 0.110 - 0.185 (幅: 0.075)
- VAモード: 類似度 0.12 - 0.24 (幅: 0.12)
- VF+VAモード: 類似度 0.115 - 0.238 (幅: 0.123)

再現率、精度



5. 実験結果と考察

設計から要件の検索結果

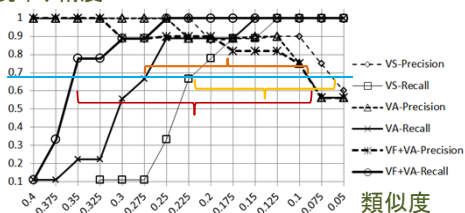


5. 実験結果と考察

再現率と精度とも0.67で類似度の範囲を評価した

- VSモード: 類似度 0.06 - 0.225 (幅: 0.165)
- VAモード: 類似度 0.08 - 0.275 (幅: 0.195)
- VF+VAモード: 類似度 0.08 - 0.355 (幅: 0.225)

再現率、精度



発表の流れ:

1. 研究の背景
2. 処理プロセスの全体像
3. SysMLによる要求・仕様記述
4. ベクトル空間検索モデルの改善
5. 実験結果と考察
6. 今後の研究方針

6. 今後の研究方針

- 結論
 - ✓VF+VAモードが安定した検索を実現できた
- 今後の研究方針
 - ◆SysMLによる記述事例を追加する
 - ◆緊密用語集合の適用戦略を考案する

ご清聴ありがとうございました
質問をお願いします