2015年度ソフトウェア工学分野の先導的研究支援事業

データマイニング手法を応用した定性的信頼性／安全性解析支援ツールの開発

広島大学大学院工学研究科

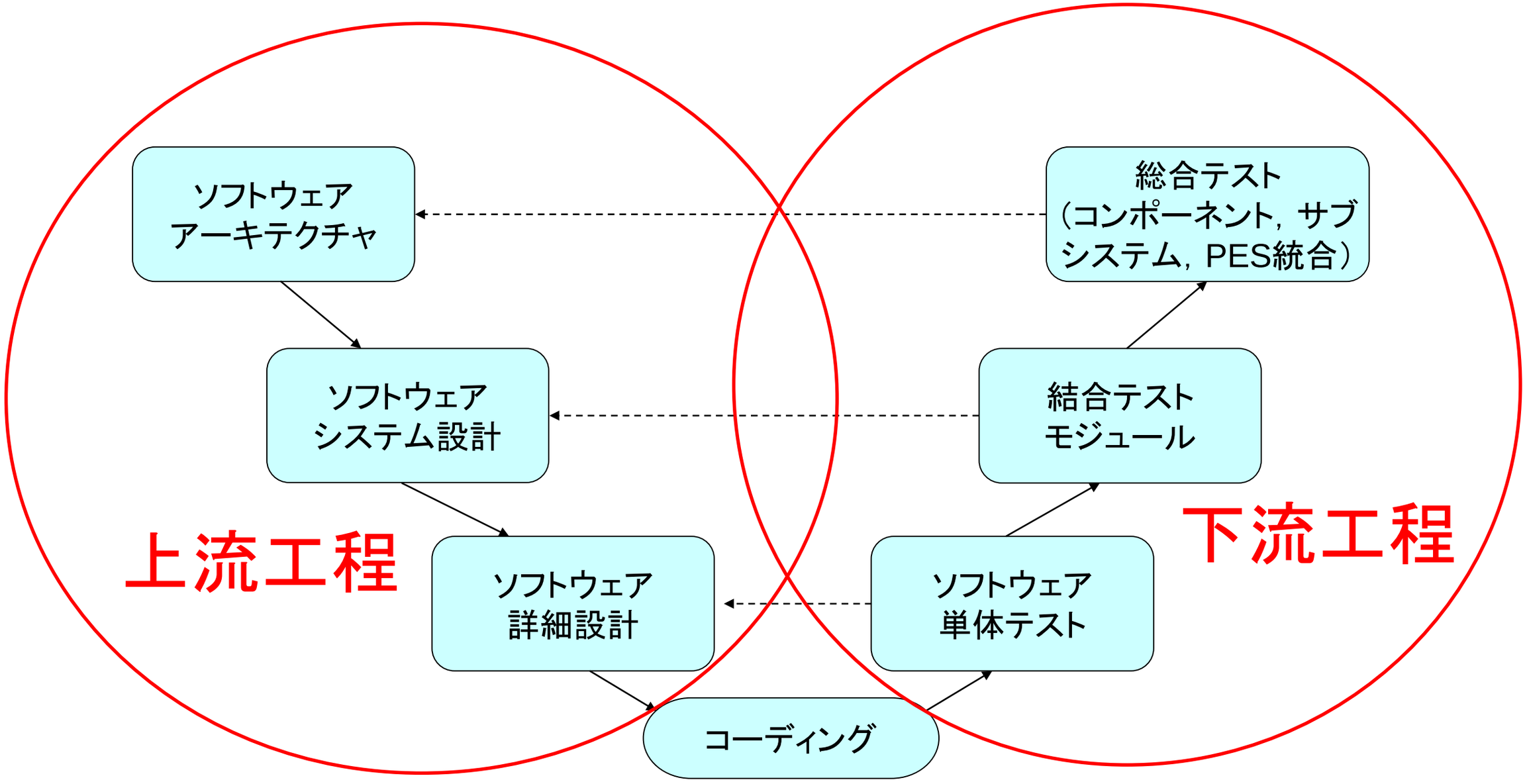
教授・土肥正（研究責任者）
准教授・岡村寛之（研究分担者）
RA・羅超 (研究分担者)

HIROSHIMA UNIVERSITY

1. 研究概要

- **背景・課題**
  - **ソフトウェアの信頼性を作り込み（確保し），かつ，信頼性を定量的に評価することは容易ではない！**
    - ✓ 多くのソフトウェアの障害は決定論的 (deterministic)
    - ✓ ソフトウェアにフォールトが含まれないことを証明することは困難
  - **定性的ソフトウェア信頼性**： ソフトウェアシステムが，定められた環境の下で規定の期間中，健全に動作する能力または度合い
  - **定量的（狭義の）ソフトウェア信頼性**： ソフトウェアシステムが，定められた環境の下で規定の期間中，健全に動作する確率（**ソフトウェア信頼度**）
  - **広義のソフトウェア信頼性**：ソフトウェアディペンダビリティ
  - **伝統的ソフトウェア工学における信頼性評価技術**： 下流工程（テスト環境）において観測されたバグ情報に基づいて，ユーザもしくはマーケットにおいてソフトウェアフォールトに起因する障害が発生しない確率を推定

ソフトウェアの信頼性/安全性技術



？

2013年度ソフトウェア工学分野の
先導的研究支援事業

HIROSHIMA UNIVERSITY

上流工程における評価が困難な理由

「上流工程における信頼性・安全性評価が困難」な理由

- 仕様書と設計情報だけから（コーディング前に）最終成果物の評価を行うため，想定される運用プロファイルや障害シナリオに基づいて主観的評価を行わざるを得ない
- 定性的モデルを定量的モデルに自動変換するため，一貫性を保証するための技術的課題
- 故障モードやエラー伝搬等の障害記述を漏れなく行うことが困難

企業における現実的な課題

- 機能安全規格による安全性分析の要請
- 安全性分析（HAZOP, FMEA, FTA）を実施するまでに至らないケースが多発
- 安全性分析による工数の大幅な増加

問題の所在

根本原因解析 (Root Cause Analysis: RCA): 問題や事象の根本的な原因や因果関係を明らかにする技術(定性的安全性技術)

- バリア分析
- 変化分析
- 特性要因図
- パレート分析
- FMEA (Failure Mode and Effect Analysis)
- FTA (Fault Tree Analysis)

問題は分析手法の選定ではなく,「想定外」の事象の考慮漏れ

- どのような事象に気をつければ良いのか？
- 障害発生シナリオを想定するための方針は何か？

研究目標

定性的信頼性 / 安全性分析を支援する新しい技術の開発

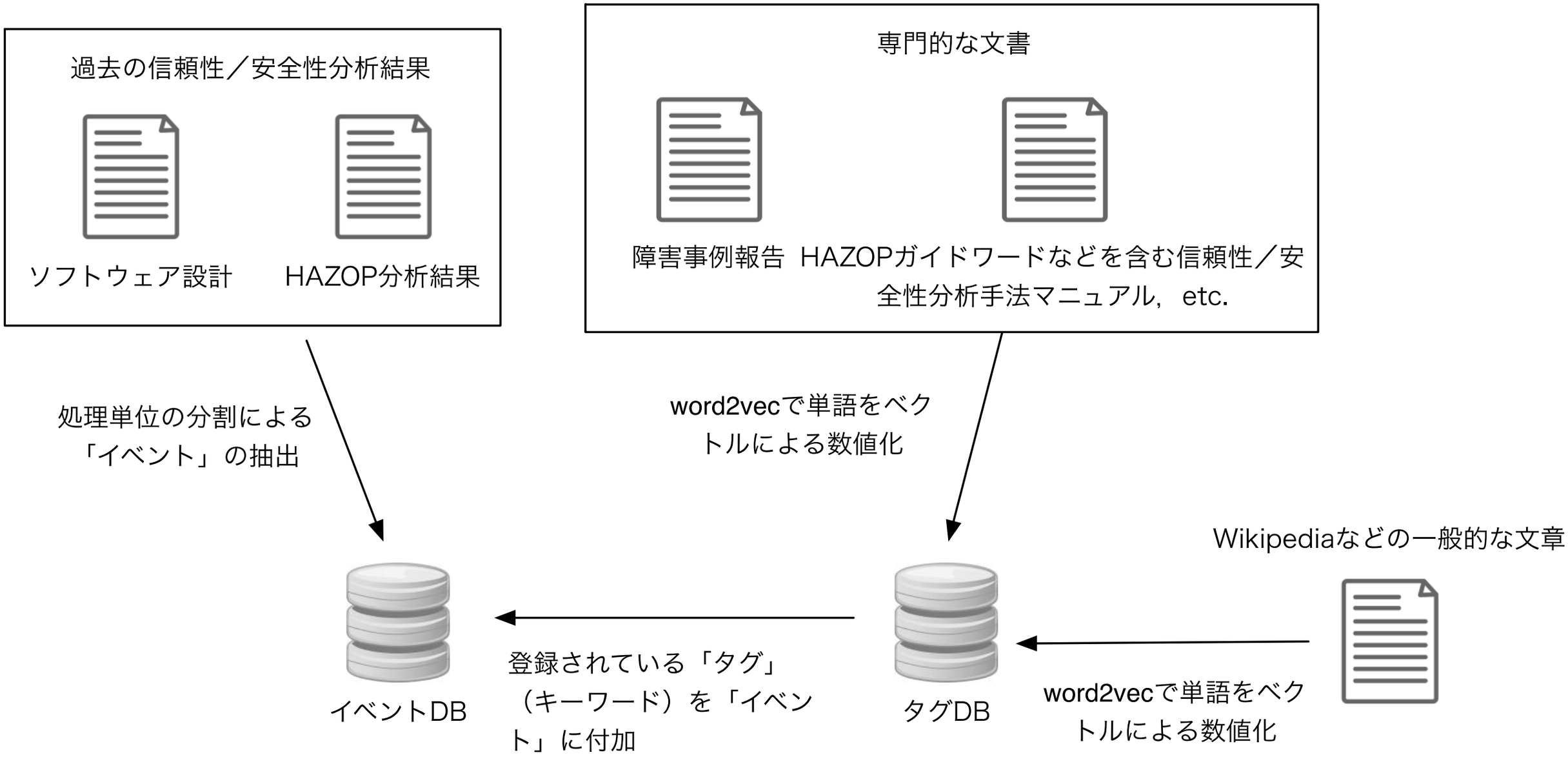
- ✓ **過去の情報（設計情報やHAZOP障害分析結果など）をデータベース上に蓄積**
- ✓ 新たな**UML/ SysML** の**設計情報と過去の情報**の相互関係（類似性）を（トピック分析）
- ✓ 新たな**UML/ SysML** の**設計情報を**安全性に寄与する重要度に従ってランキングするためのアルゴリズムを開発

開発現場において分析を支援するツールの開発

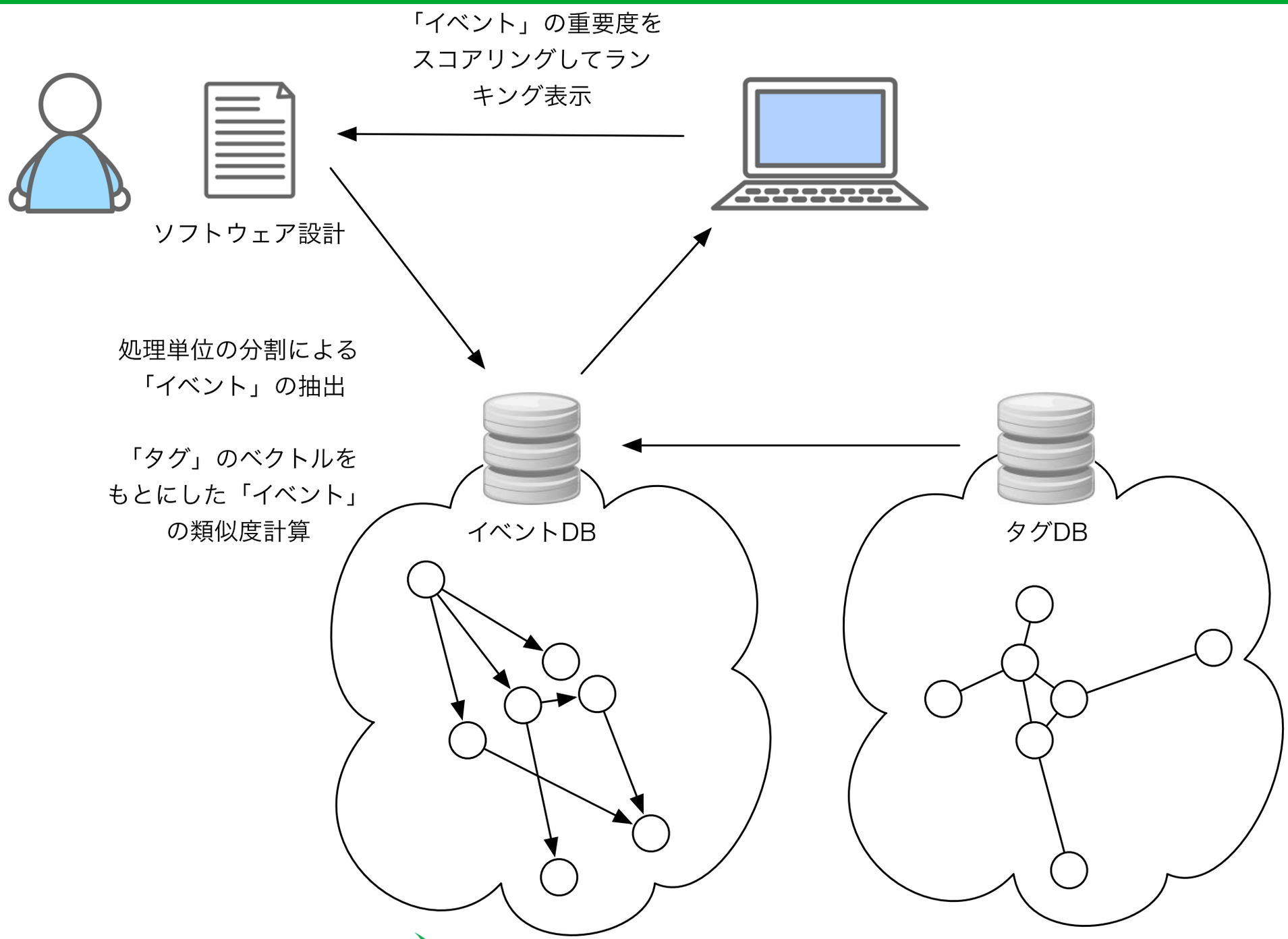
- ✓ 新たな設計と過去のHAZOP分析との紐付けができるため分析経験の少い分析者でも, 故障事象の考慮漏れが起きない
- ✓ 重要度によりランキングされているため，重要度の低い部分の分析を化することで安全性分析に対するコストを減らす

2. 研究成果

開発ツールの概要 - 1/2

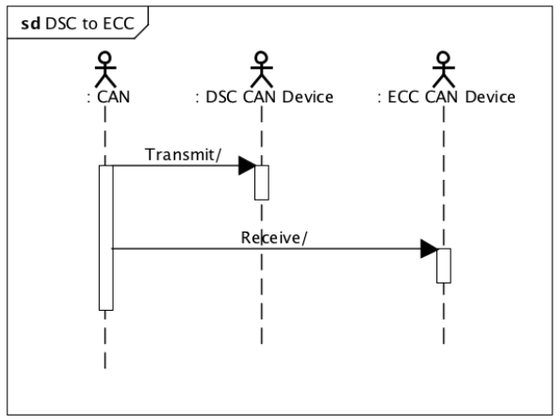


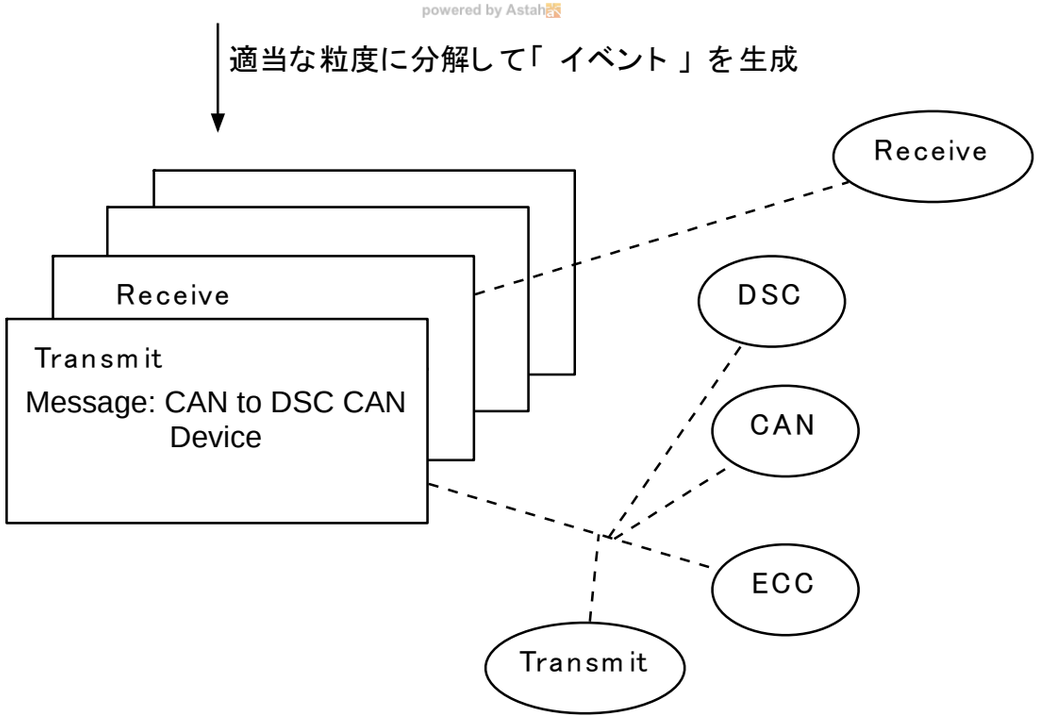
開発ツールの概要 – 2/2


「イベント」の重要度を
スコアリングしてラン
キング表示
ソフトウェア設計
処理単位の分割による
「イベント」の抽出
「タグ」のベクトルを
もとにした「イベント」
の類似度計算
イベントDB
タグDB

検索手法の調査・開発（研究課題1）

- ソフトウェア設計書／安全性分析などの各種ドキュメントをどのようなデータ構造で扱うか
  - イベント：ある特定された粒度の文章（文書の1段落，HAZOP分析における一つの障害シナリオ，
    シーケンス上の一つのメッセージなど）
    - テキスト文章
    - イベント間のリンク：直接関連するイベント
- タグ：イベントのテキスト文章から抽出されるキーワード
  - 共起による数値ベクトル化（word2vec）





類似イベント検索

- タグ間の類似度：COS類似度

$$\mathrm{sim}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$$

- イベント間の類似度：タグの出現頻度による重み付きベクトルのCOS類似度

$$\boldsymbol{v} = \sum_k w_k x_k$$

  - $\boldsymbol{v}$:
  - w_k: k
  - x_k: k

スコアリング手法の調査・開発（研究課題2）

- タグをどのようしてベクトル数値化するか
  - word2vec
    - 2013年 Google の研究所
    - 「同じ文脈で利用される単語は、同じ意味を持つ」と言う仮説（共起）
    - 自然言語で記述された文章（学習データ）から単語の「意味」をベクトル化する
    - ニューラルネットワークを利用した数値化
      → ニューラルネットワークの中間層の値によるベクトル化
      （深層学習と同様）
    - king – man + woman = queen
- MeCab
  - 日本語形態素解析システム
  - word2vec の学習データ生成（わかち書き）に利用

スコアリング手法の調査・開発（研究課題2）

- 類似度から重要度をどのようにして算出するか
  - 「過去の安全（HAZOP）分析結果」に5段階スコア
  - 「新しいイベント」と「過去の安全分析結果」の類似度を算出

    - $\mathrm{score}(i)$: シーケンスのイベント i の重要度
    - $\mathrm{sim}(i,k)$: シーケンスのイベント i と HAZOP 分析結果のイベント k の類似度
    - $\mathrm{score}(k)$: HAZOP 分析結果のイベント k の重要度（HAZOP スコア）
    - $I(\cdot)$: 指標関数

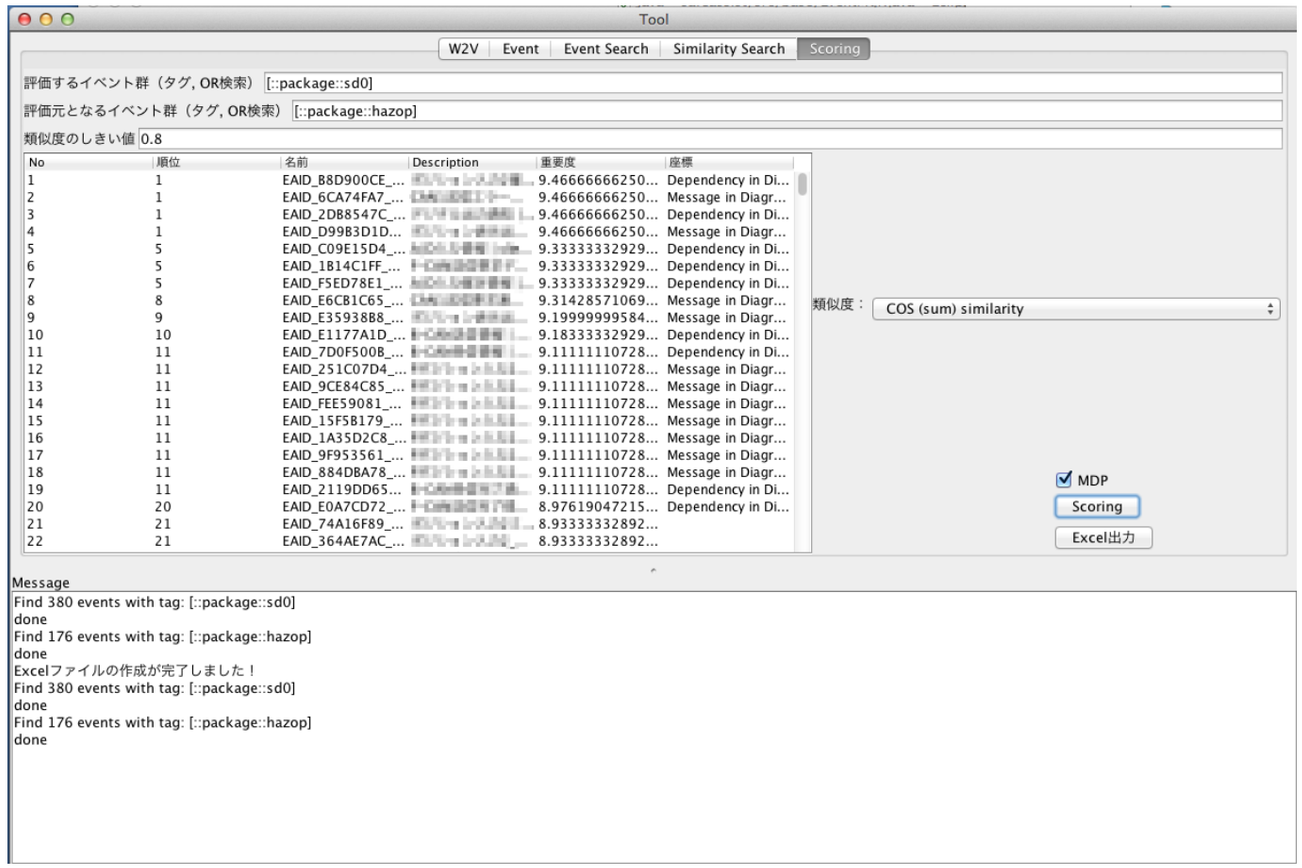
      類似度が閾値 θ 以上のHAZOPスコアの平均

$$\mathrm{score}(i) = \frac{\sum_{k \in H} I(\mathrm{sim}(i,k) > \theta)\mathrm{score}(k)}{\sum_{k \in H} I(\mathrm{sim}(i,k) > \theta)}$$

+ 上記のスコアをもとにした MDP（マルコフ決定過程）によるスコア

ツール開発（研究課題4）

- Java + MeCab + SQLデータベース による開発
- UMLシーケンスに対する重要度評価
  （1）word2vec データの作成
  （2）シーケンスおよび安全性分析結果の登録
  （3）シーケンスの重要度算出



有効性検証（研究課題5）

- 実務におけるシーケンスおよび安全分析結果からシーケンスの重要度を算出
  - タグDB（Wikipedia，IPAが発行する電子テキスト，etc.）
  - シーケンス図 SD0, SD1
  - 安全性分析（HAZOP分析）結果
- 事前に分析者によって5段階評価が付与

要素/機能ID	要素/機	ガイドワード	発生する逸脱	対象コン	他コンポーネン	考えられる原因	対策有無	対処方策	安全方策ID	重要度
1-1-1	Pポジシ	Omission (処理が存在しない)	対象がデータであるため、処理としてみることはない	-	-	-	-	-	-	1
	Pポジシ	Commission (不要な処理が存在する)	対象がデータであるため、処理としてみることはない	-	-	-	-	-	-	1
	Pポジシ	Early (処理の提供が早い)	押されるよりも早く押されるということはありえないので、逸	-	-	-	-	-	-	1
	Pポジシ	Late (処理の提供が遅い)	PポジションSWがすぐに効かない	なし	SWの反応が遅い	レジスタ故障 ・レジスタの変化が遅い RAM異常 ・一時的に化ける	有り	RAMをサイクリックに更新する	SSR 4-1-10-1	4
	Pポジシ	Value (値の間違い)	PポジションSW以外が出力される	なし	OFFなのにON、ONなのにOFFとなる	レジスタ故障 ・レジスタが固着 RAM異常 ・RAMのビットが固着	有り	RAMをサイクリックに更新する	SSR 4-1-10-1	4

有効性検証（研究課題5）

SD0

No	順位	名前	Description	重要度	座標
1	1	EAID_7D0F5		5	Dependency in Diagram
2	1	EAID_B8D9C		5	Dependency in Diagram
3	1	EAID_E0A7C		5	Dependency in Diagram
4	1	EAID_E3593		5	Message in Diagram P
5	1	EAID_251C0		5	Message in Diagram P
6	1	EAID_C09E1		5	Dependency in Diagram
7	1	EAID_40F35		5	Message in Diagram P
8	1	EAID_1B14C		5	Dependency in Diagram
9	1	EAID_6CA74		5	Message in Diagram P
10	1	EAID_2DB85		5	Dependency in Diagram
11	1	EAID_8C22A		5	Message in Diagram P
12	1	EAID_9CE84		5	Message in Diagram P
13	1	EAID_D99B3		5	Message in Diagram P
14	1	EAID_E2FBA			am P
15	1	EAID_91653			am P
16	1	EAID_FEE59			am P
17	1	EAID_BC1E7			am P
18	1	EAID_15F5B			am P
19	1	EAID_8164F			am P
20	1	EAID_1A35D			am P
21	1	EAID_9F953			am P
22	1	EAID_E6CB			am P
23	1	EAID_F5ED7			iagram
24	1	EAID_B6716			am P
25	1	EAID_E1177			iagram
26	1	EAID_884DE			Message in Diagram P
27	1	EAID_3EE79		5	Message in Diagram P
28	1	EAID_F53A6		5	Message in Diagram P
29	1	EAID_2119D		5	Dependency in Diagram
30	1	EAID_64ABE		5	Message in Diagram P
31	31	EAID_9E849		4.2	
32	31	EAID_E6360		4.2	
33	31	EAID_5A57E		4.2	Comment in Diagram
34	31	EAID_CD3E5		4.2	

380 イベント中

重要度 5
30 イベント

重要度 0
185 イベント

SD1

No	順位	名前	Description	重要度	座標
1	1	EAID_D99B3		5	Message in Diagram P
2	1	EAID_64ABE		5	Message in Diagram P
3	1	EAID_1B14C		5	Dependency in Diagram
4	1	EAID_B6716		5	Message in Diagram P
5	1	EAID_251C0		5	Message in Diagram P
6	1	EAID_2DB85		5	Dependency in Diagram
7	1	EAID_6CA74		5	Message in Diagram P
8	1	EAID_FEE59		5	Message in Diagram P
9	1	EAID_9CE84		5	Message in Diagram P
10	1	EAID_8164F		5	Message in Diagram P
11	1	EAID_E1177		5	Dependency in Diagram
12	1	EAID_7D0F5		5	Dependency in Diagram
13	1	EAID_8C22A			Diagram P
14	1	EAID_E0A7C			in Diagram
15	1	EAID_C09E1			in Diagram
16	1	EAID_15F5B			Diagram P
17	1	EAID_9F953			Diagram P
18	1	EAID_2119D			in Diagram
19	1	EAID_E3593			Diagram P
20	1	EAID_3EFFE			Diagram P
21	1	EAID_F5ED7			in Diagram
22	1	EAID_B8D9C			in Diagram
23	1	EAID_8DF5E			Diagram P
24	24	EAID_1ECF4			
25	24	EAID_F6057			
26	24	EAID_BE90E		4.2	
27	24	EAID_C8D78		4.2	Message in Diagram P
28	24	EAID_35930		4.2	
29	24	EAID_2AAB5		4.2	

380 イベント中

重要度 5
23 イベント

重要度 0
186 イベント

3. 結論と今後の課題

- 成果活用見込み
  - 結果の検証
    - 過去のHAZOP分析と新たな設計が同一アプリケーションドメインである場合，安全性分析者にとって比較的納得のいくスコアリング／フィルタリングができている
      （前回の報告での質疑応答に対する回答）
- 改善点
  - MeCabの辞書精度が大きく影響：今回はMeCab標準の辞書＋Wikipediaの見出しを辞書としたが，安全性分析固有の言葉を間違って分解
    →用語辞書の整備・オープン化
  - 結果表示：「イベント」がシーケンス図上のどこに位置するのかがわかりにくい
    →UML/SysMLツールとの融合

Contd.

- 研究成果の発表，投稿，引用など
  - ツール（主として検索エンジン部分）のオープンソース化
  - GUIインタフェース整備＋付加機能（UMLツールとの統合など）で製品化