

非機能要件抽出手法の構築

60180132

宮崎智己

研究の背景

.....

- 近年のシステム開発では、委託開発が主流
- ベンダ側は開発するシステムに対する顧客の要件を正確に把握する必要がある
 - RFP（調達仕様書）や要件定義書から必要な要件を把握

要件	概要
機能要件	システムが満たすべき 機能に関する要件のこと
非機能要件	機能以外の要件のこと システムの性能やセキュリティなど

非機能要件（NFR） [1]

.....

要件	概要
可用性	システムの継続性に関する項目
性能・拡張性	システムの性能や拡張性に関する項目
運用・保守性	運用・保守作業に関する項目
移行性	新システムへの移行に関する項目
セキュリティ	セキュリティに関する項目
システム環境・エコロジー	消費電力など自然環境に関する項目

[1]非機能要件グレード, <https://www.ipa.go.jp/sec/softwareengineering/reports/20100416.html>

NFR特定の課題

.....

- システムの短納期化が進んでいるため、早急にNFRを特定する必要がある
- NFRを手で特定するのは大変
 - NFRは非常に多くの項目が対象
 - RFPや要件定義書は多くの自然言語で記述
 - 多いものでは200ページ以上にも

**RFPや要件定義書に含まれている
非機能要件を特定することは非常にコストがかかる**

業務のシステム最適化を進め、平成21年3月には第Ⅰ期間発分として労働基準行政情報システム（以下「基準システム」という。）と労災行政情報管理システム（以下「労災システム」という。）との共通基盤部分が稼働している。

また、第Ⅱ期間発分（業務処理の集中化等）として平成22年12月に基準システム

先行研究

- NFRを自動分類する手法が提案されている[2]
 - 教師あり学習を用いて分類する手法

データの暗号化をすること	セキュリティ
応答時間は2秒以内	性能性
⋮	⋮

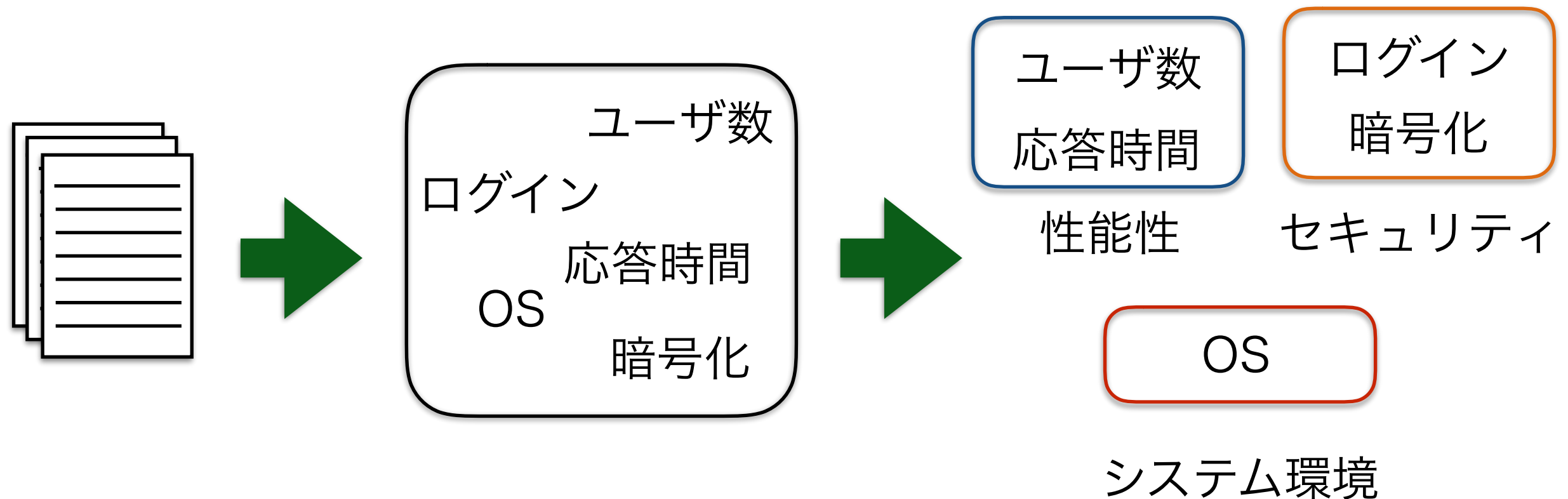
**学習データを作成するコストがかかる
そのままでは新しいデータに対して適用できない**

SVM等

[2]Wen Zhang, Ye Yang, Qing Wang, Fengdi ShuAn Empirical Study on Classification of Non-Functional Requirements, Proceedings of the 23rd International Conference on Software Engineering & Knowledge Engineering (SEKE'2011)

先行研究

- 教師なし学習を用いた手法が提案[3]
 - NFRごとに関連のある単語の辞書を作成
 - 階層クラスタリングを行い、単語のクラスタを作っている
 - 対象とする文と作成した辞書の類似度で分類する

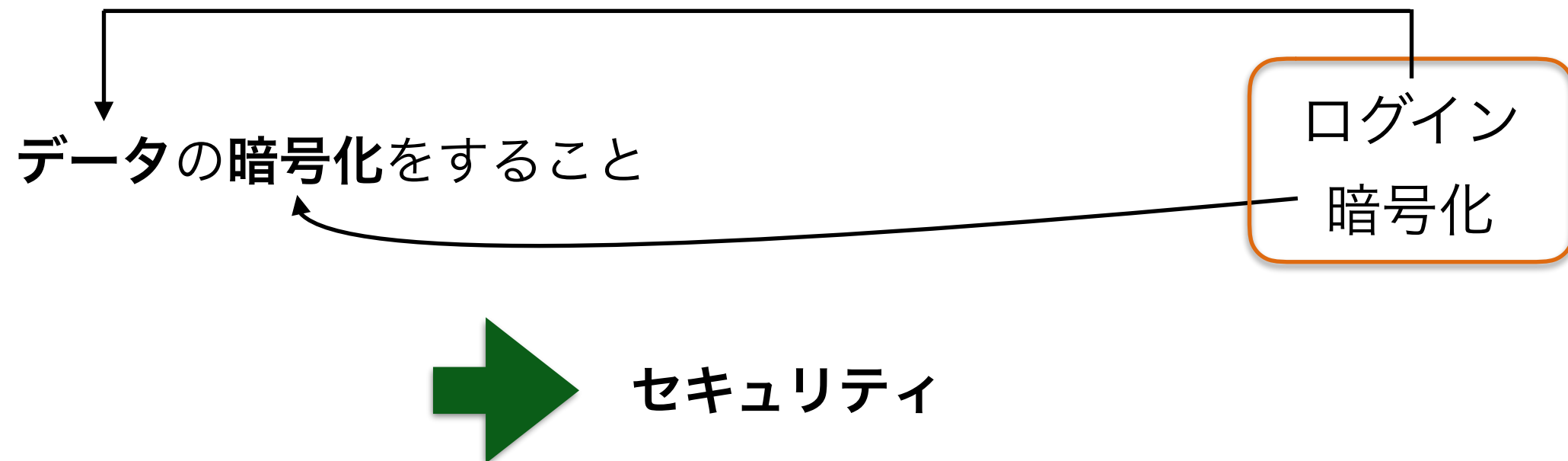


[3]Anas Mahmoud, An information theoretic approach for extracting and tracing non-functional requirements, In Proceedings IEEE 23rd International Requirements Engineering Conference (RE'15), pp. 36-- 45, 2015

先行研究

.....

- 教師なし学習を用いた手法が提案[4]
 - NFRごとに関連のある単語の辞書を作成
 - 階層クラスタリングを行い、単語のクラスタを作っている
 - 対象とする文と作成した辞書の類似度で分類する



[4]Anas Mahmoud, An information theoretic approach for extracting and tracing non-functional requirements, In Proceedings IEEE 23rd International Requirements Engineering Conference (RE'15), pp. 36-- 45, 2015

先行研究の課題

- 先行研究では， 粒度の大きい分類をしている
 - 性能性， セキュリティ， 運用・保守性. . .
- 利用者は， 先行研究の手法を用いて分類を行った後，
改めて分類する必要がある

研究の目的とアプローチ

- 目的

- 先行研究と同程度，もしくはそれ以上の精度で，より粒度の細かいNFRの分類を実現すること

- アプローチ

- 考え中

現在のTODO

- 先行研究の手法で粒度の細かい分類をやってみる
 - 細かく分類するために，何が問題なのかを明らかにするため
- 実装は完了，あとは実験
 - 類似度計算にNormalized Google Distance (NGD) を利用
 - Googleの検索用APIを使っているが，一日のアクセス制限があり計算が終わらない．．．

補足資料

先行研究の辞書作成手順

1. 単語の関連度の計算

- Normalized Google Distance (NGD) [1]を用いる
 - 同じWebページに含まれている単語は関連がある

$$NGD(w_1, w_2) = \frac{\max\{\log(D_1), \log(D_2)\} - \log(|D_1 \cap D_2|)}{\log(|D|) - \min\{\log(D_1), \log(D_2)\}}$$

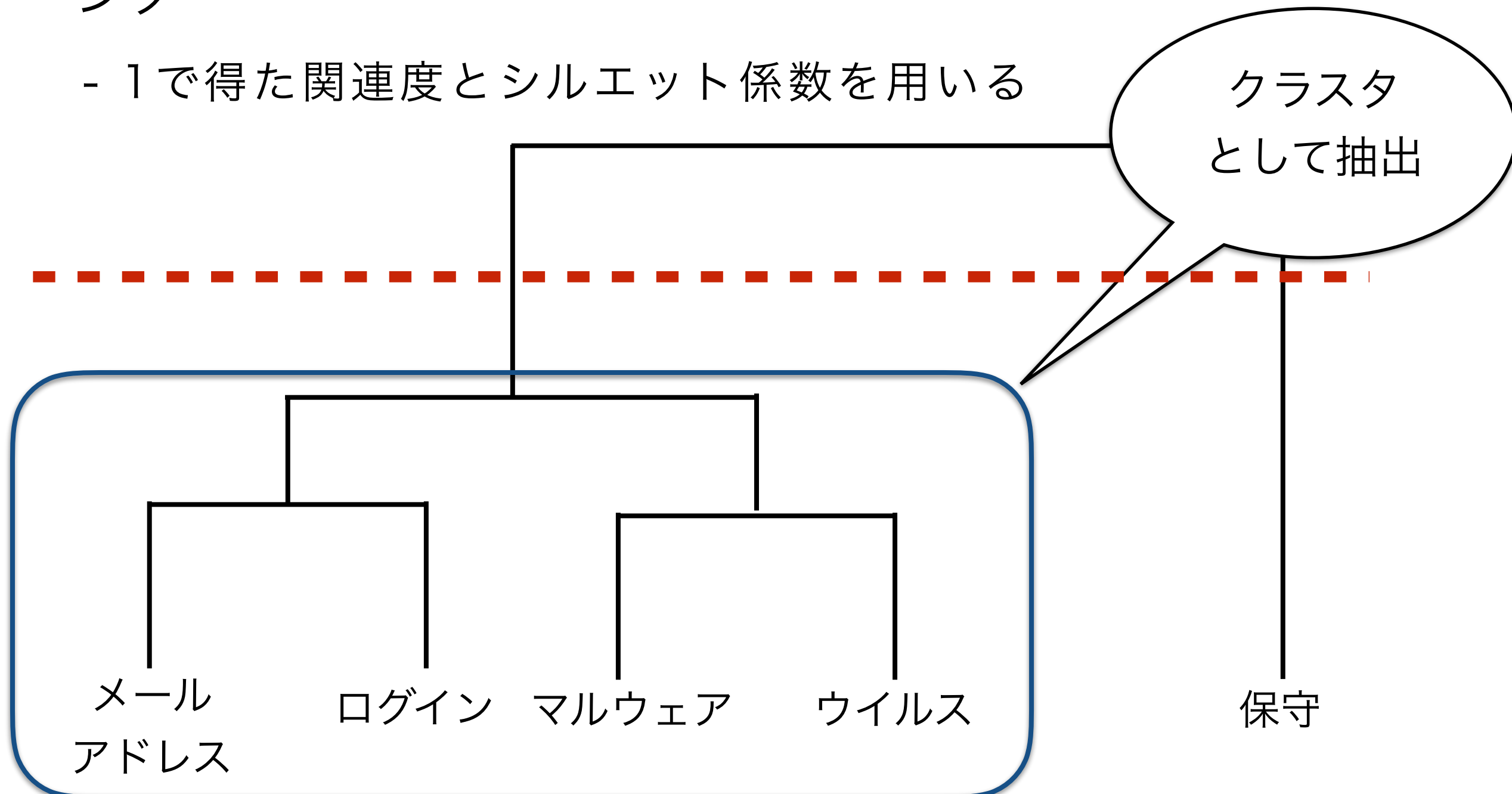
D : インデックスされているページの総数 D_1 : キーワード w_1 を含むページ数
 D_2 : キーワード w_2 を含むページ数 D_3 : キーワード w_1 と w_2 を含むページ数

- 例) 「ログイン」と「メールアドレス」 (全ページ1000百万件の場合)
 - ログイン: 611百万件・メールアドレス: 120百万件
 - ログイン&メールアドレス: 18.8百万件

$$\text{関連度} = \frac{\log 611 - \log 18.8}{\log 1000 - \log 120} = 0.34$$

2. 単語のクラスタリング

- 凝集型階層クラスタリングを用いて単語をクラスタリング
 - 1で得た関連度とシルエット係数を用いる



3. クラスタに対してNFRを割当てる

- クラスタに含まれる単語と非機能要件の各カテゴリに含まれる単語の関連度をNGDで調べ、NGDが最も高いカテゴリをそのクラスタのラベルとする

$$P(C, L) = \frac{1}{N} \sum_{i=1}^N \max_{0 \leq j \leq M} NGD_{wiki}(w_i, l_j),$$

w_i : クラスタCに含まれる単語

N : クラスタCに含まれる単語の総数

l_j : 非機能要件Lに含まれる単語

M : 非機能要件Lに含まれる単語の総数

