

ドメインオントロジーを用いた 要求獲得法の提案

海谷 治彦 (信州大学)

佐伯 元司 (東京工業大学)

2006年4月21日

目次

- 背景と目的
- オントロジーを用いた意味処理の枠組み
- 要求分析の手順
- ケーススタディの報告

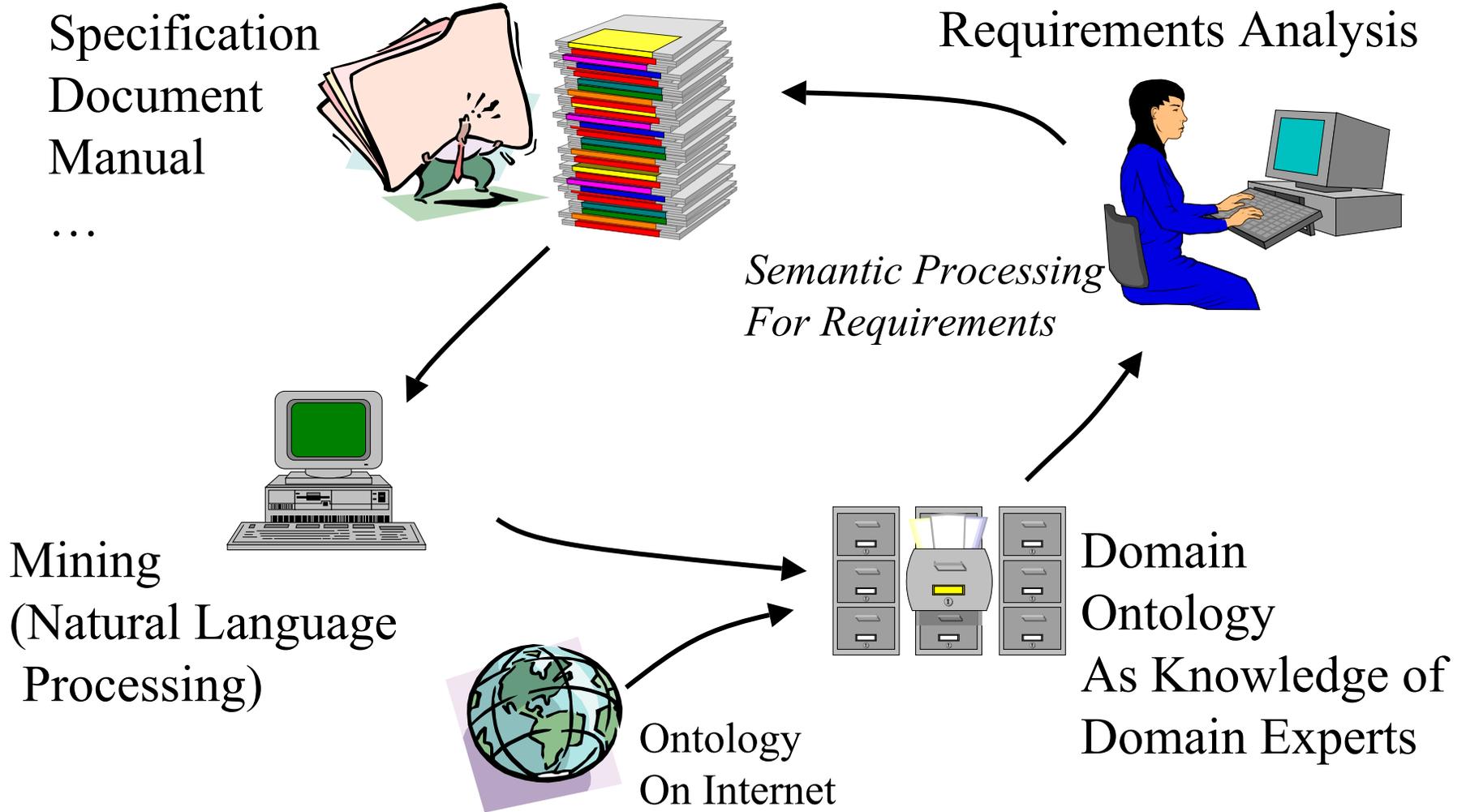
背景

- 情報システムの構築にはシステムを利用する予定の業務(ドメイン)の知識が必須.
- ドメイン知識を業務の専門家だけから直接に得るのはコストや時間がかかりすぎる.
- 「書き物」になっているドメイン知識を用いて「意味」まで考慮した要求分析を(半)自動的に行うのは困難.

プロジェクト全体の目的と方針

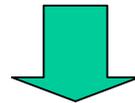
- 計算機処理可能なドメイン知識(表現)を用いて、意味を考慮した要求分析を行う。
 - 表現としてオントロジーを用いる。
- 可能な限り意味処理を自動的に行う。
 - 再現率の重視 (疑わしきは罰する！)
- ドメイン知識は既存のリソースから可能な限り自動抽出する。
 - 既存の情報システムに関する文書
 - 既存の辞書やシソーラス本発表では本件には触れない。

ゴールの青写真



本発表の論点

- 要求仕様書の機械的な分析支援.
- 字面の情報で可能な限り意味まで踏み込んだ分析をする方法.
- 特殊な記述様式(半形式言語や厳格な形式仕様)の仕様書を想定しない.



- 低コストな方法で意味まで考慮した意味処理を実現する！ (軽量化意味処理)

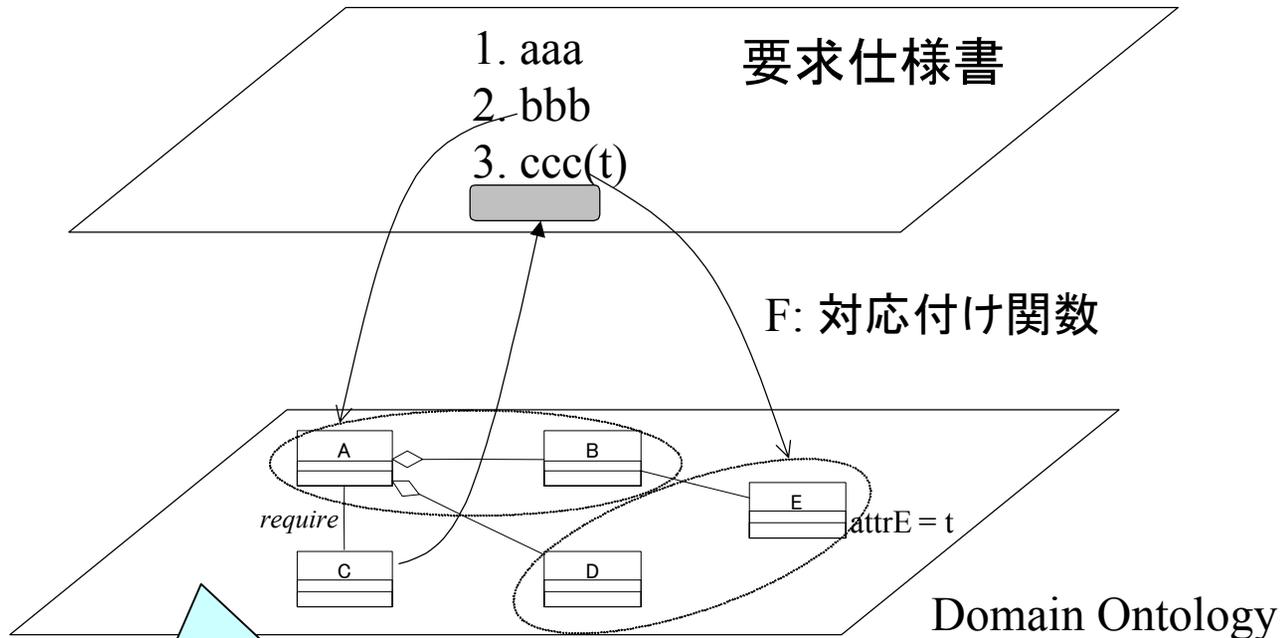
要求分析で考慮すべき特性

- 矛盾
 - 矛盾した記述で設計・実装したら大変.
- 欠落事項
 - 後段での要求追加は回避したい.
- コレクトネス
 - 客の要望に合致してないと, そもそも意味無し.
- 非曖昧性
 - 後段での誤解・誤実装の原因となる.

オントロジー: 意味解釈の基盤

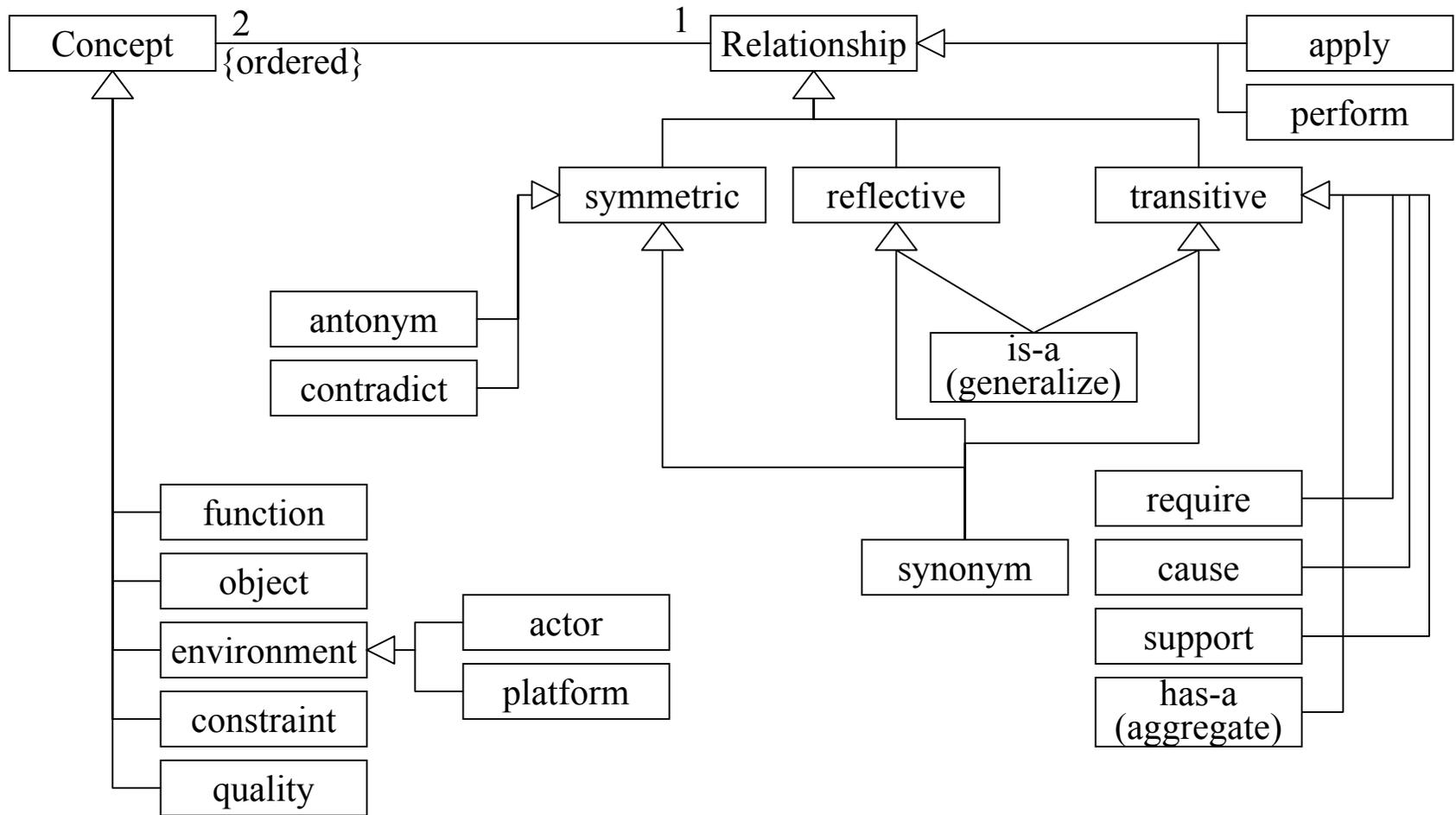
- 有名な定義
 - formal explicit specification of shared conceptualization
- 本研究では,
 - 要求項目の意味付けに使える構造
 - 意味付けをする構造上で計算をする機構として使う.
- オントロジー ⇒ シソーラス + 推論規則

要求仕様書とオントロジの関係



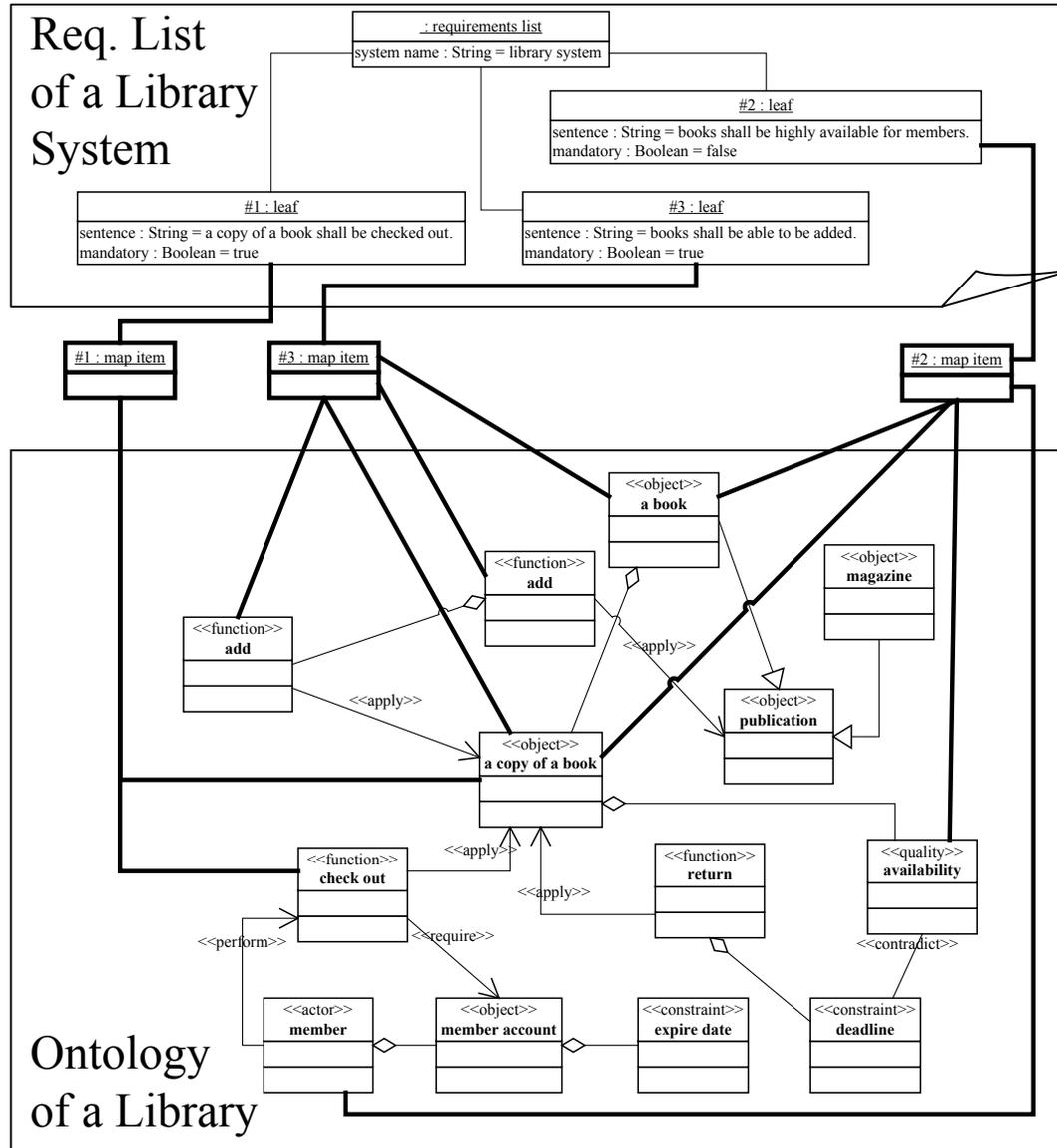
この上で各種の計算(推論)をすることで意味処理を代行する.

オントロジーのデータ構造



要求分析に都合がよい型を導入.

例



意味処理の方針

- 要求仕様書の「文全体」の意味には立ち入らない。
 - 例えば否定文の否定部分は扱わない。
- 文を構成する「単語」レベルの意味に注目する。
- 注目する意味は、前述のように、1. 矛盾 2. 抜け 3. ニーズにあってない 4. 曖昧 の4点。
- 意味的に疑わしい部分は可能な限り抽出。
 - 誤判定ならば、分析者が却下すればよい。

要求分析手順の明確化

- 軽量化意味処理のアイデアに基づき実際に分析を進める手順が必要。
 - アイディアだけでは実際に分析者が分析を行うのは困難.
- 全自動とはいかないので、自動化できる部分と人がやる部分の切り分けの明確化が必要.
- 要はCASEツールの仕様が必要.

手順の概要

1. 要求仕様を単文(要求項目)に整形
2. 要求項目とオントロジを対応付け
3. メトリクス(後述)に基づき要求仕様の現状の「意味的な良さ」を判定
 - 十分に良ければ分析終了
4. 「良くない」と判定された部分を修正.
 - 良くするためにどんな意味を追加/削除すればよいかを提示.
 - 提示に基づき要求仕様を変更
5. 2.に戻る.

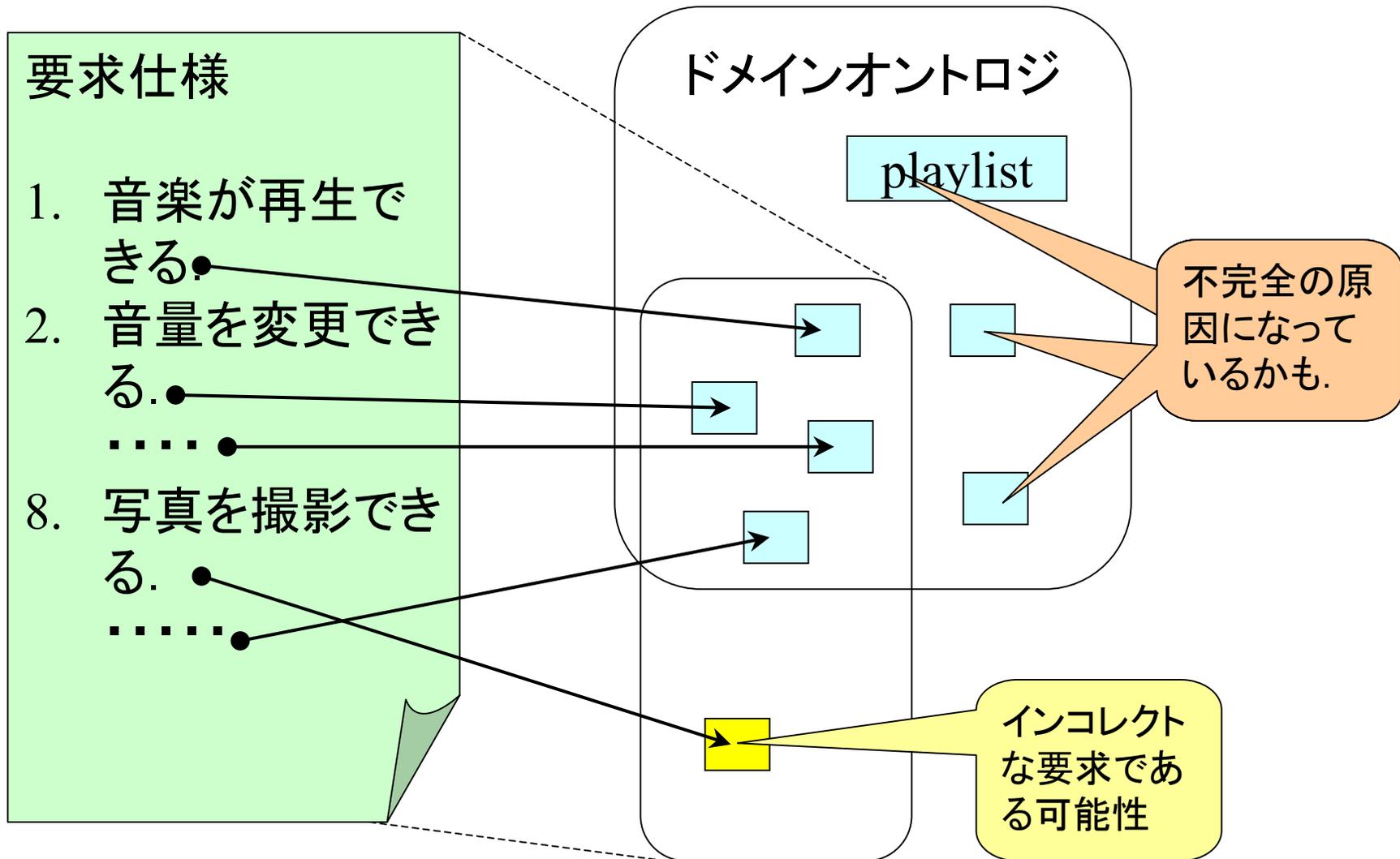
青字の部分は自動化, 他は人手が必要.

メトリクス

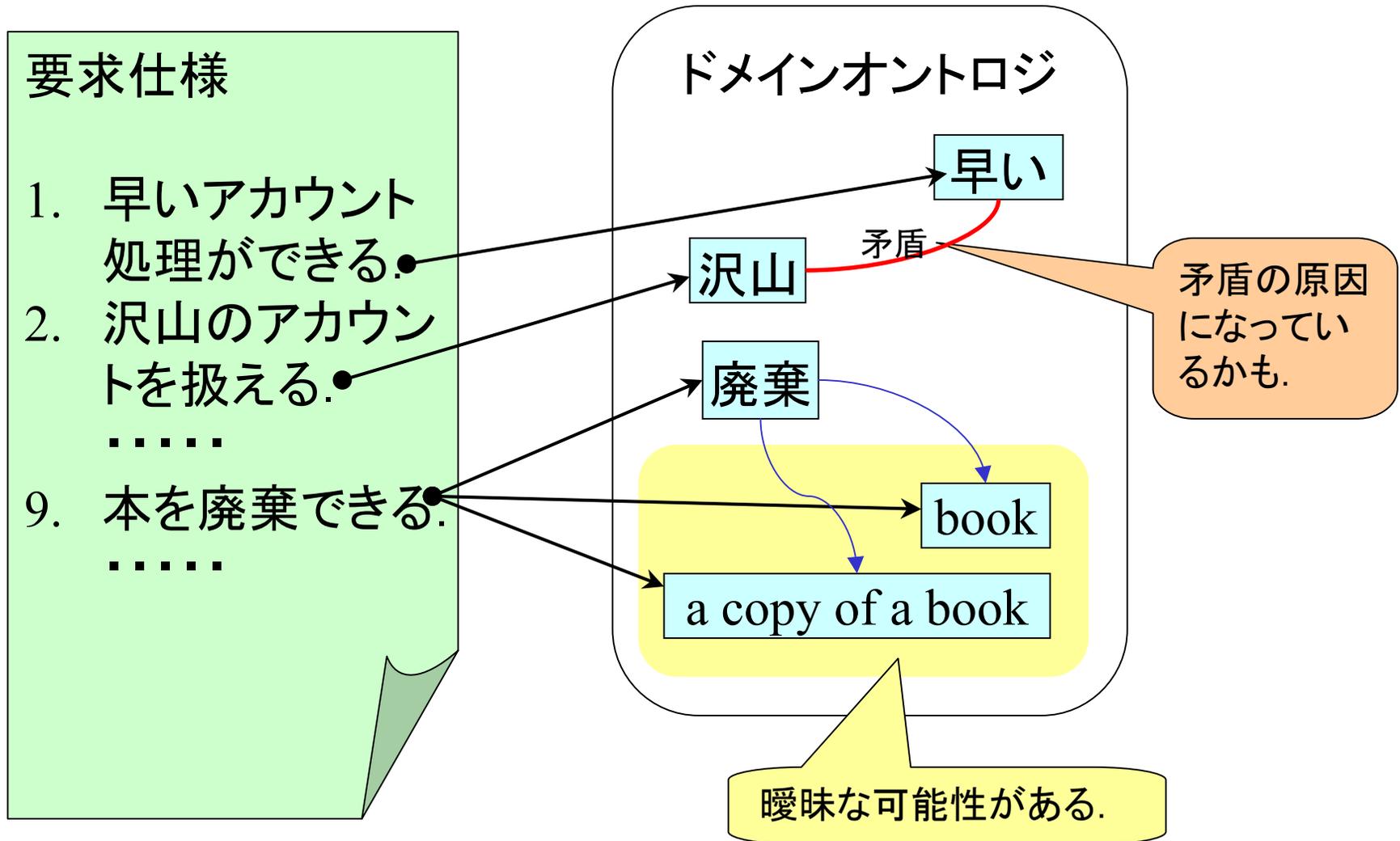
要求仕様書の「良さ」を判定するための数値

- コレクトネス:
 - 要求仕様が当該ドメインの意味要素に包含されている割合。
- 完全性:
 - 当該ドメインに含まれるべき意味が包含されている割合。
- 無矛盾性:
 - 当該ドメイン上で相互に矛盾する概念が含まれている割合。
- 非曖昧性:
 - 相互に関連が少ない概念と対応付かない要求項目の割合

メトリクスの直感的な説明 1/2



メトリクスの直感的な説明 2/2



各特性の改善法

- コレクトネスが低ければ
 - 当該ドメインに相応しくない概念は自動的に抽出できる.
 - それを削除するように要求文を人が書き換える.
- 完全性が低ければ
 - 既存の概念が必要とするにもかかわらず要求仕様で言及されていない概念は自動抽出可能.
 - それを追加するように要求文を書き換える.
- 無矛盾性が低ければ
 - どの部分がどんな相矛盾する概念を含むかは自動抽出可能.
 - それを排除するように要求文を書き換える.
- 曖昧ならば
 - どの文がどんな相互に関係ない概念と関連付いているかは自動的に指摘可能.
 - それに基づき要求文を書き直す.

比較実験

- 目的:
 - 前述の分析手順が要求分析の有効なガイドとなっているかを確認.
 - 要は手順が与えられたほうがマシンに要求分析が行われるか否かを確認.
- 問題
 - PC上で動作する音楽プレイヤー
- 被験者
 - 学部学生 S1, S2 (ガイドあり) S3 (なし)

結果の概要

手順に従った被験者

S1	初期	Step 1	Step 2
要求項目数	11	24	39
写像概念数	14	21	35
コレクトネス(%)	91	96	97
完全性(%)	12	29	46
無矛盾性(%)	-	100	100
曖昧性(%)	100	50	49

S2	初期	Step 1	Step 2
要求項目数	11	25	25
写像概念数	13	26	26
コレクトネス(%)	90	100	100
完全性(%)	11	22	22
無矛盾性(%)	100	100	100
曖昧性(%)	80	100	100

自由にやらせた被験者

S3	初期	Step 1
要求項目数	6	16
写像概念数	14	40
コレクトネス(%)	83	100
完全性(%)	12	35
無矛盾性(%)	100	98
曖昧性(%)	100	100

考察

ガイドされたほうが……

- 個々の要求文を単純に書く傾向がある.
 - 概念数/文: $S1=0.9$ $S2=1.1$ $< S3=2.5$
- 概念と文との対応付けは分析者の主観にまかせても、それほど問題はないようだ.
- メトリクスで改善を促進するのはそれなりに成功しているようだ.
- ガイドが無い被験者(S3)もオントロジを幅優先探索して改善を模索していた. ガイドも基本的にはこの方針であるため、不自然なものではないだろう.

まとめと今後の課題

- 軽量化意味処理に基づく要求獲得手順を定義.
- 同手順がそれなりに妥当であることを比較実験から確認
- 意味処理のルールを改善, 拡張
 - 「意味的誤り」の検出に関する再現率を向上させるために.
 - 精度はまあほどほどでよい.
- 手順をガイドするツールの完成