

オントロジーに基づく明細書の記述

産業技術総合研究所

情報技術研究部門

橋田 浩一

2004-07-17

概要

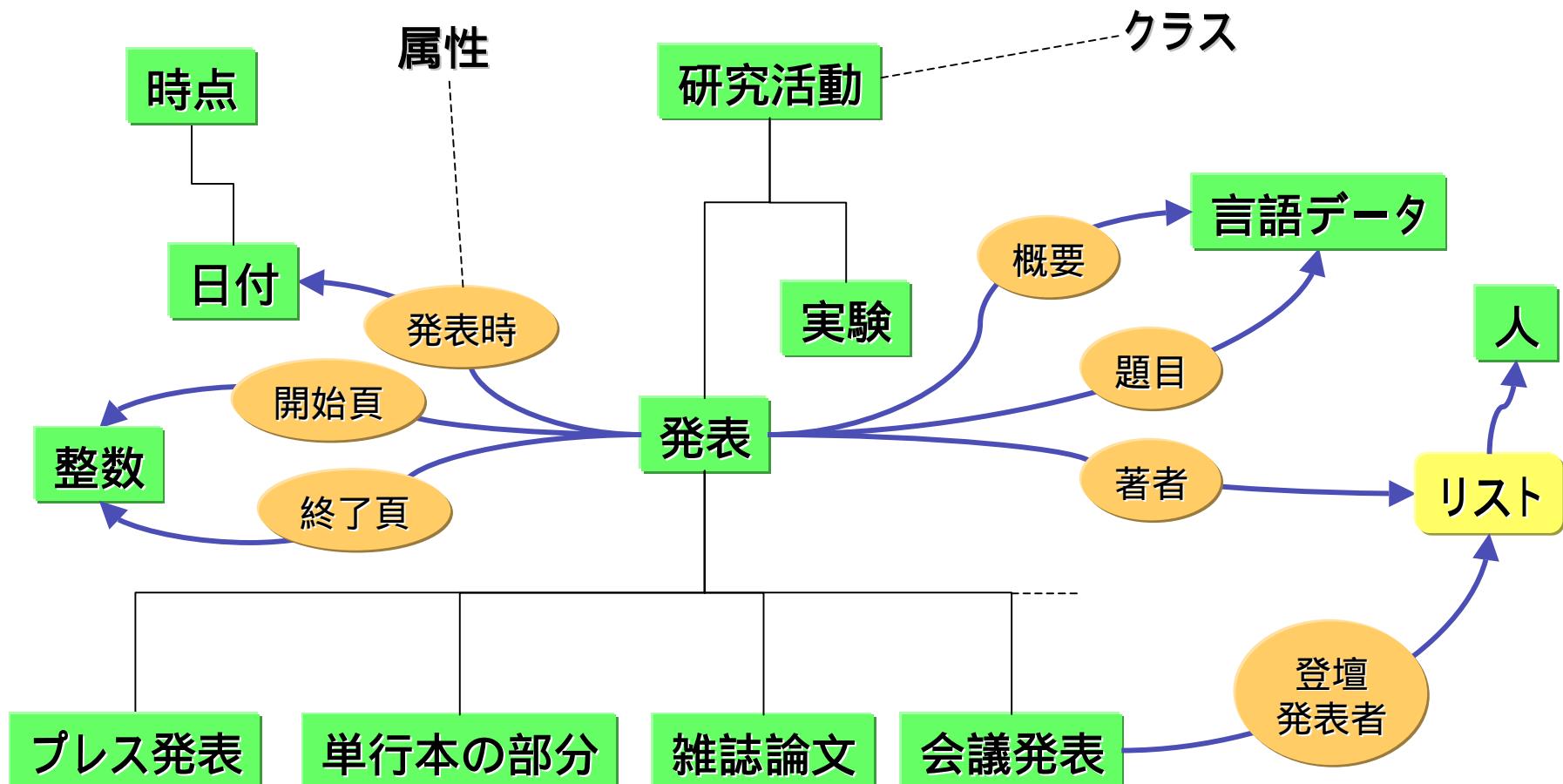
- 特許の明細書を含む文書の意味構造を一般的にグラフとして扱いたい。
- 定型的な構造はオントロジーに基づくERグラフとして記述。
- 不定型的な構造はオントロジーに基づかないERグラフとして記述。
- 明細書(特に請求項)で定型的/不定型な部分は?

オントロジー

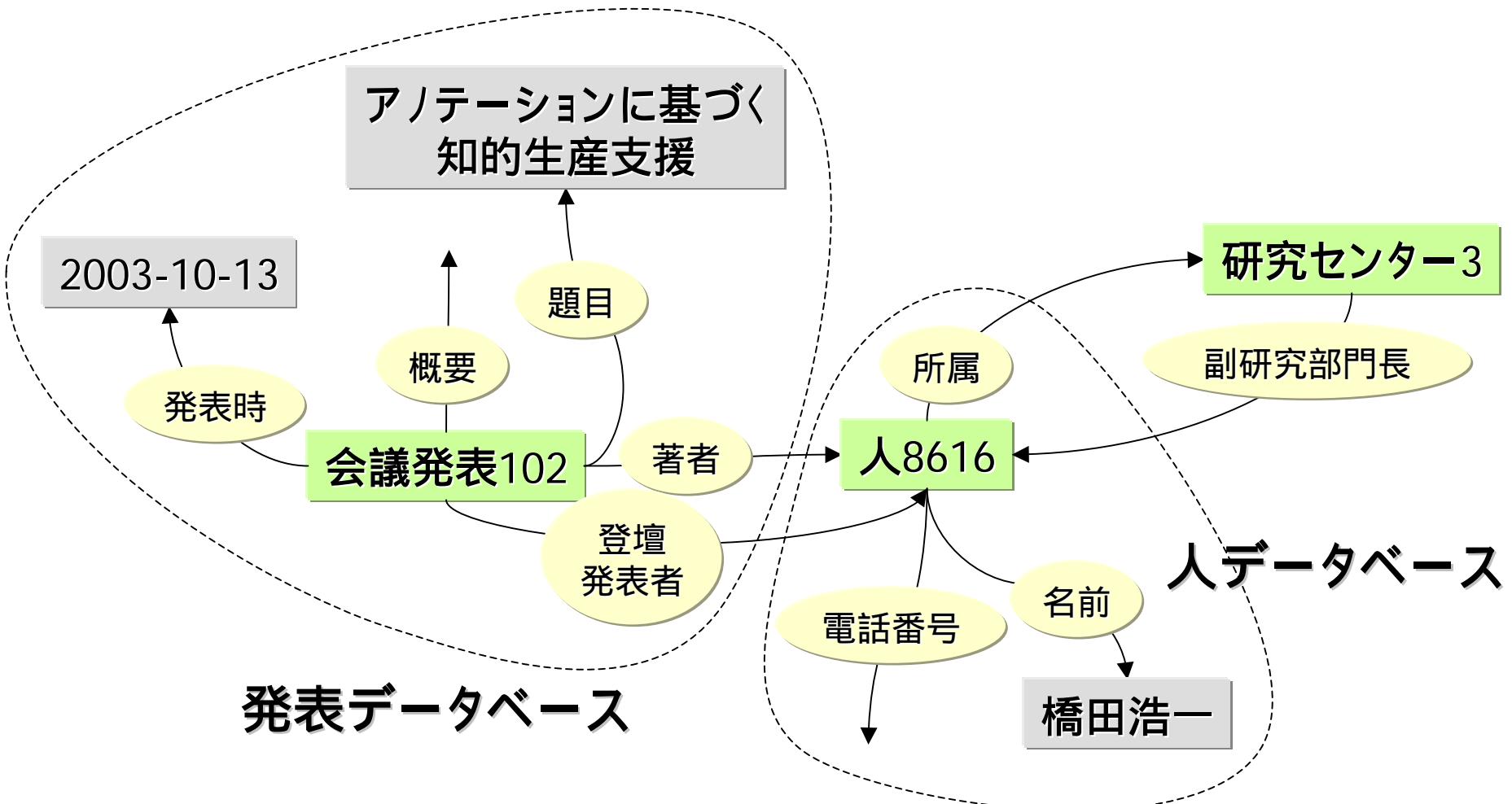
概念体系 = データベースのスキーマ

- クラス(class;概念) = 具体例(instance)の集合
例: 「情報技術研究部門」は「研究ユニット」クラスの具体例
- 分類(taxonomy)のラティス
 - クラスの上位下位(包含)関係
 - 上位クラスは集合として下位クラスを含む
例: 「発表」は「雑誌論文」の上位概念
- 属性(property)
 - クラスの各具体例が持つ性質(2項関係)
例: 「橋田」の「所属ユニット」属性の値は「情報技術研究部門」である
 - 各項の型(要素、集合、リスト)と要素のクラス
 - 上位クラスからの継承

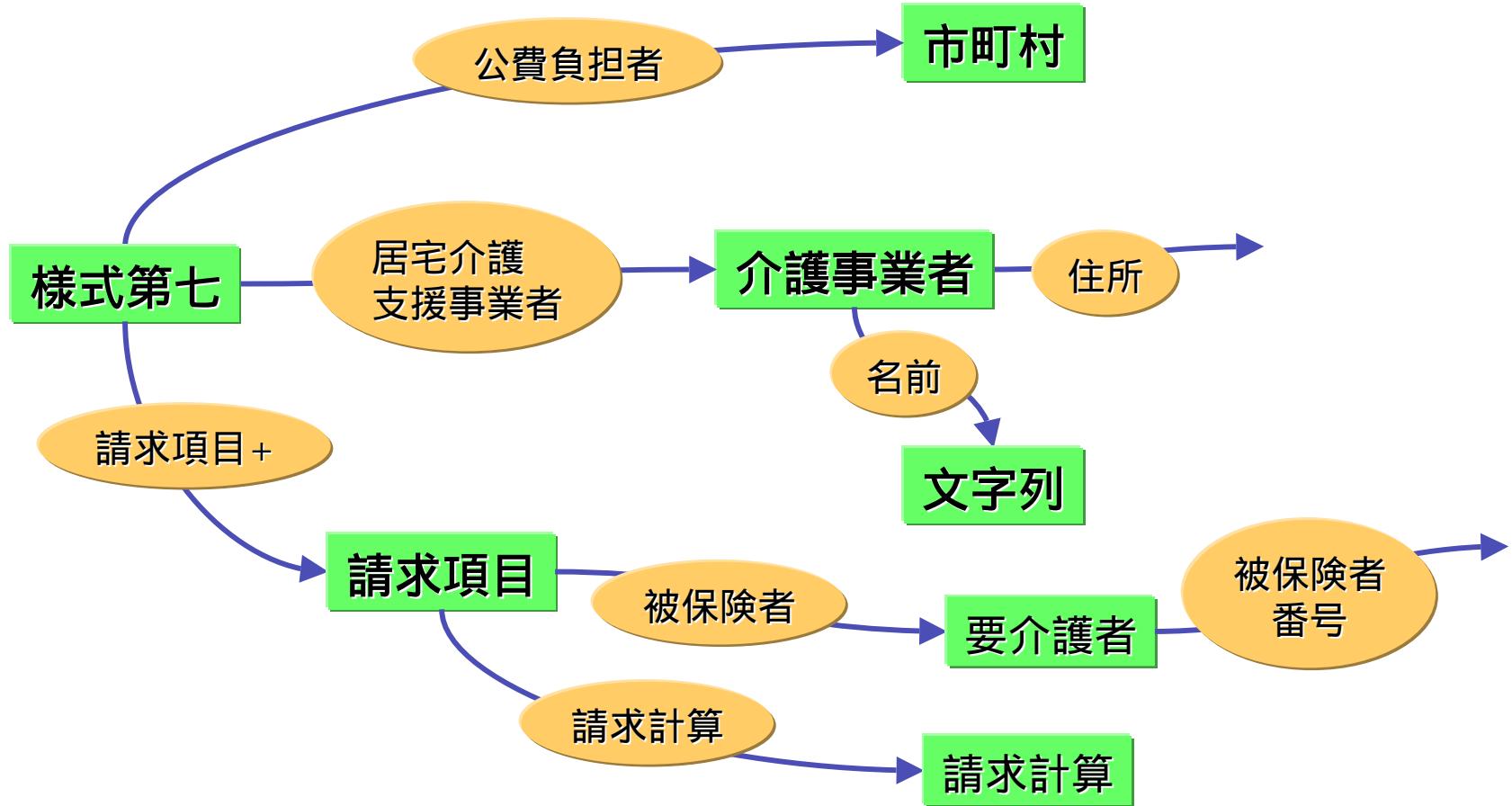
オントロジーの例: 成果発表DB



オントロジーに基づく知的コンテンツ … 実体関係(ER)モデル



文書のオントロジー



様式第七（附則第二条関係）

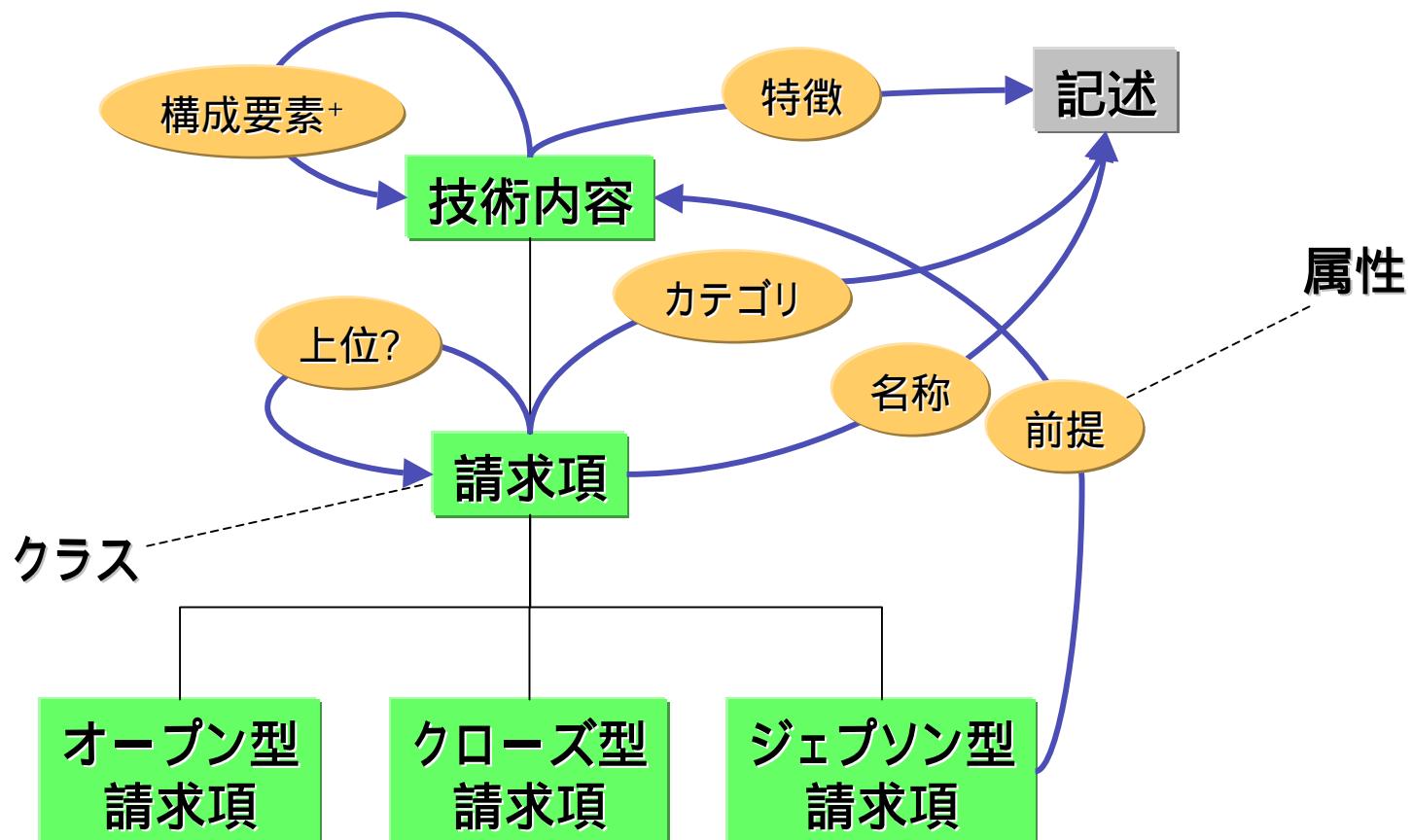
居宅介護支援介護給付費明細書

公費負担者番号								平成			年		月分					
													保険者番号					

居宅介護 支援事業者	事業所 番号								所在地	〒		一		
	事業所 名称	連絡先 電話番号												
		単位数単 価		▲			(円/単位)							

項目番号	被保険者													請求計算								
	被保険者番号	(フリガナ) 氏名														サービスコード						
公費受給者番号																						
生年月日	1. 明治 2. 大正 3. 昭和	性別	1. 男 2. 女													単位数						
	年 月 日		平成		年		月		日	から												
要介護 状態区分	要支援・1・2・3・4・5	認定 有効期間	平成		年		月		日	まで	請求金額											
居宅サービス計画作成依頼届出年月日															平成		年		月		日	
被保険者番号	(フリガナ) 氏名														サービスコード							
公費受給者番号																						
生年月日	1. 明治 2. 大正 3. 昭和	性別	1. 男 2. 女													単位数						
	年 月 日		平成		年		月		日	から												
要介護 状態区分	要支援・1・2・3・4・5	認定 有効期間	平成		年		月		日	まで	請求金額											
居宅サービス計画作成依頼届出年月日															平成		年		月		日	
被保険者番号	(フリガナ) 氏名														サービスコード							
公費受給者番号																						
生年月日	1. 明治 2. 大正 3. 昭和	性別	1. 男 2. 女													単位数						
	年 月 日		平成		年		月		日	から												
要介護 状態区分	要支援・1・2・3・4・5	認定 有効期間	平成		年		月		日	まで	請求金額											
居宅サービス計画作成依頼届出年月日															平成		年		月		日	

PCMLに対応するオントロジー

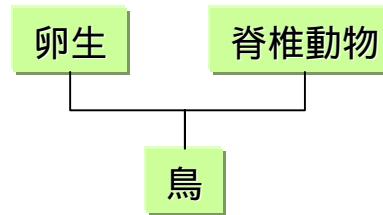


XMLとオントロジー

- XMLはシンタクス(統語論)を規定
 - XMLに基づく記述はシンタクスを表わす木
- オントロジーはセマンティクスを規定
 - オントロジーに基づく記述は意味を表わすグラフ
- 意味構造は木でなくグラフ
 - オントロジーに基づいて記述すべし
- オントロジーに基づく記述法は意味に即して規定されるゆえにXMLによる記述法より原理が明確

XMLにできなくて オントロジーにできること

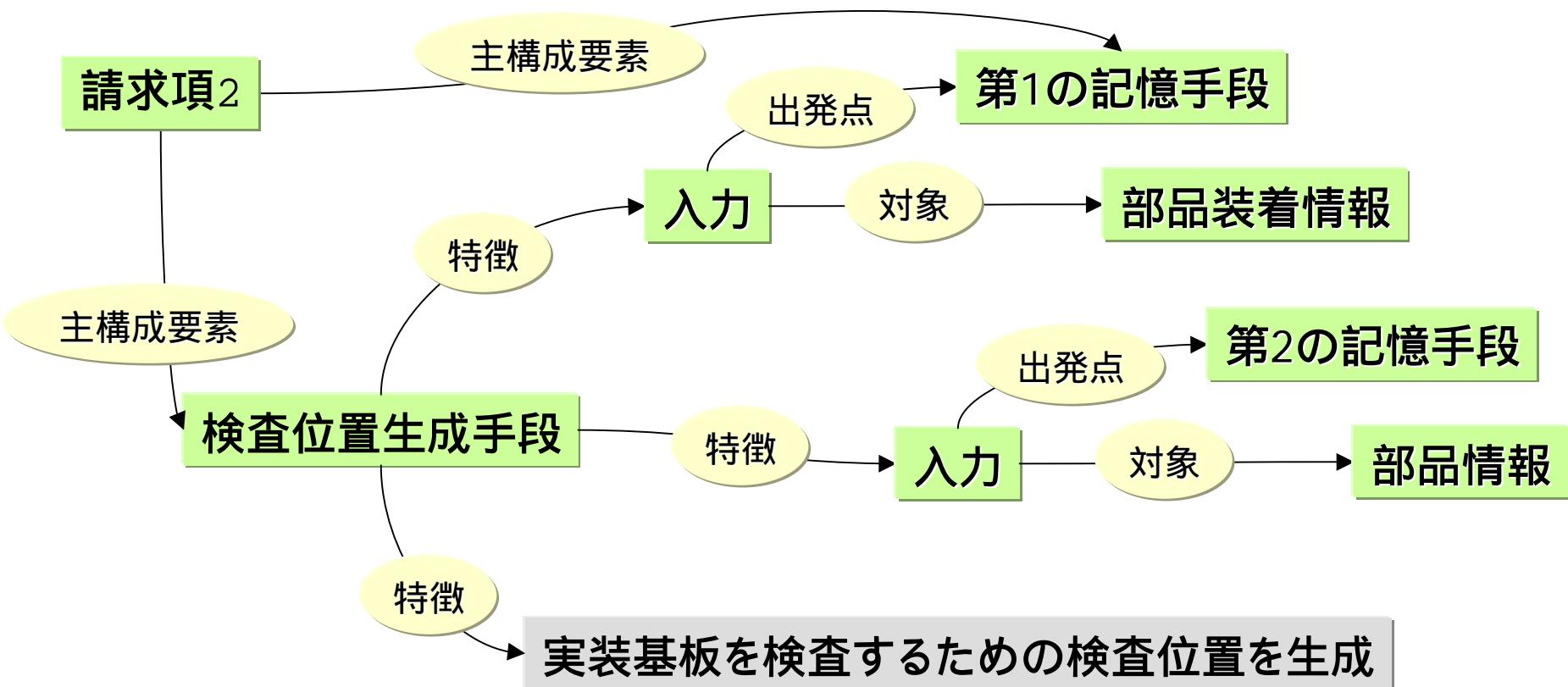
- グラフの表現
 - XMLで直接表現できるのは木だけ
 - XMLでは参照(idref、URI reference等)の対象のクラスを指定できない。
- 多重継承
 - 複数個の親クラス
 - XML Schema等では单一継承しかできない。



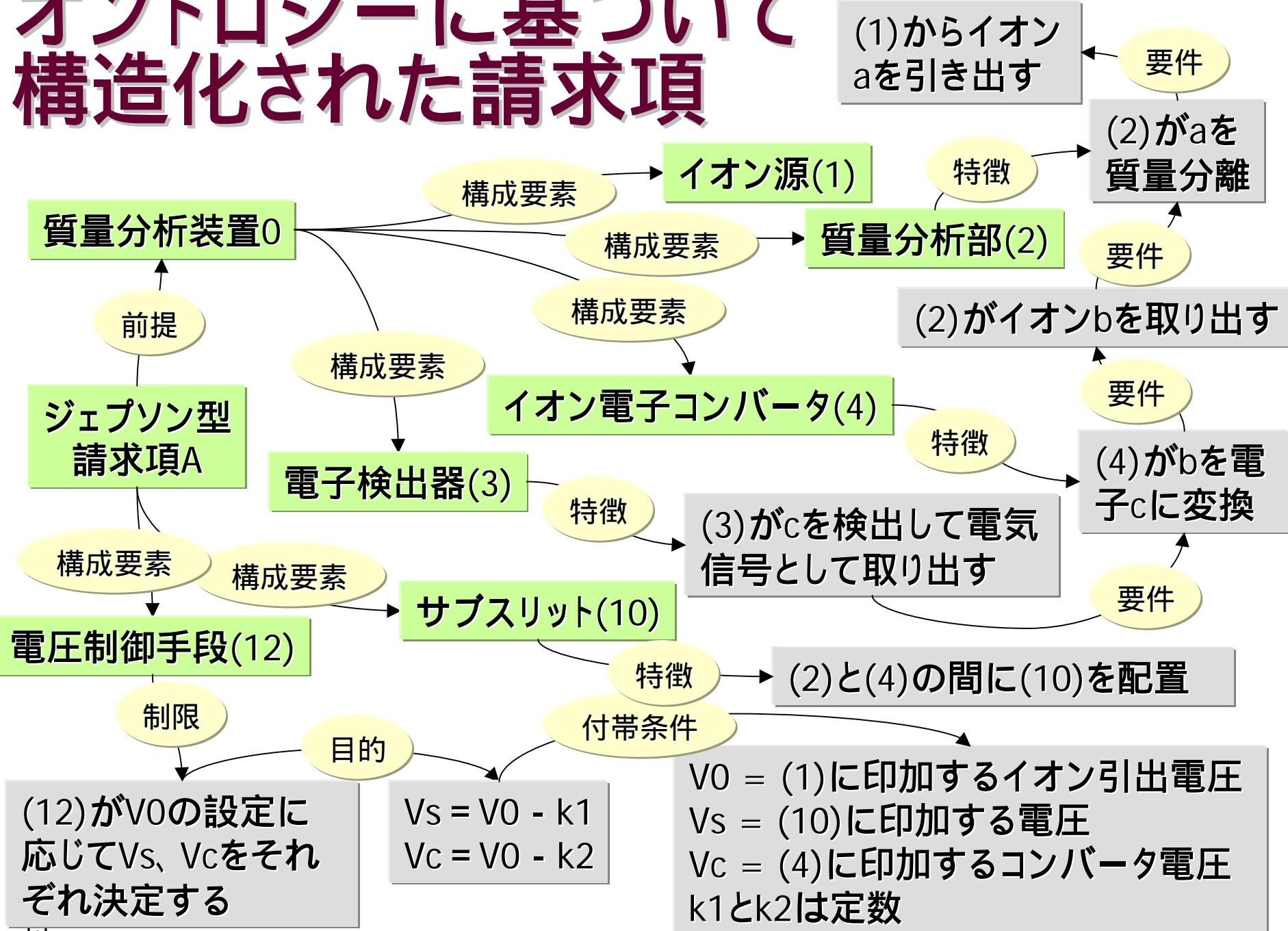
XML (PCML)に基づく記述

```
<構成要素 特徴性="主特徴部">
  <定義情報>
    <相対的記述>
      <入出力関係 方向="入力">
        <源構成要素名>第1の記憶手段</源構成要素名>
        <入出力名称>部品装着情報</入出力名称>
      </入出力関係>
      <入出力関係 方向="入力">
        <源構成要素名>第2の記憶手段</源構成要素名>
        <入出力名称>部品情報</入出力名称>
      </入出力関係>
    </相対的記述>
    <絶対的記述>
      <自己特性>実装基板を検査するための検査位置を生成</自己特性>
    </絶対的記述>
  </定義情報>
</構成要素>
<構成要素名>検査位置生成手段</構成要素名>
```

オントロジーに基づく記述



オントロジーに基づいて構造化された請求項



まとめ

- オントロジーの利用
 - あり … 特定分野のコンテンツ
 - なし … 他のコンテンツ
- (特許などの)特定分野での意味構造化を普及させるにはオントロジーの標準化と共有が重要
- PCMLの本体はDTDではなく概念辞書であるべし
- (部分的に)オントロジーに基づいて特許の明細書を含むさまざまな文書をオーサリングする作業を支援する一般的なソフトウェアツールが必要 … 作成中