

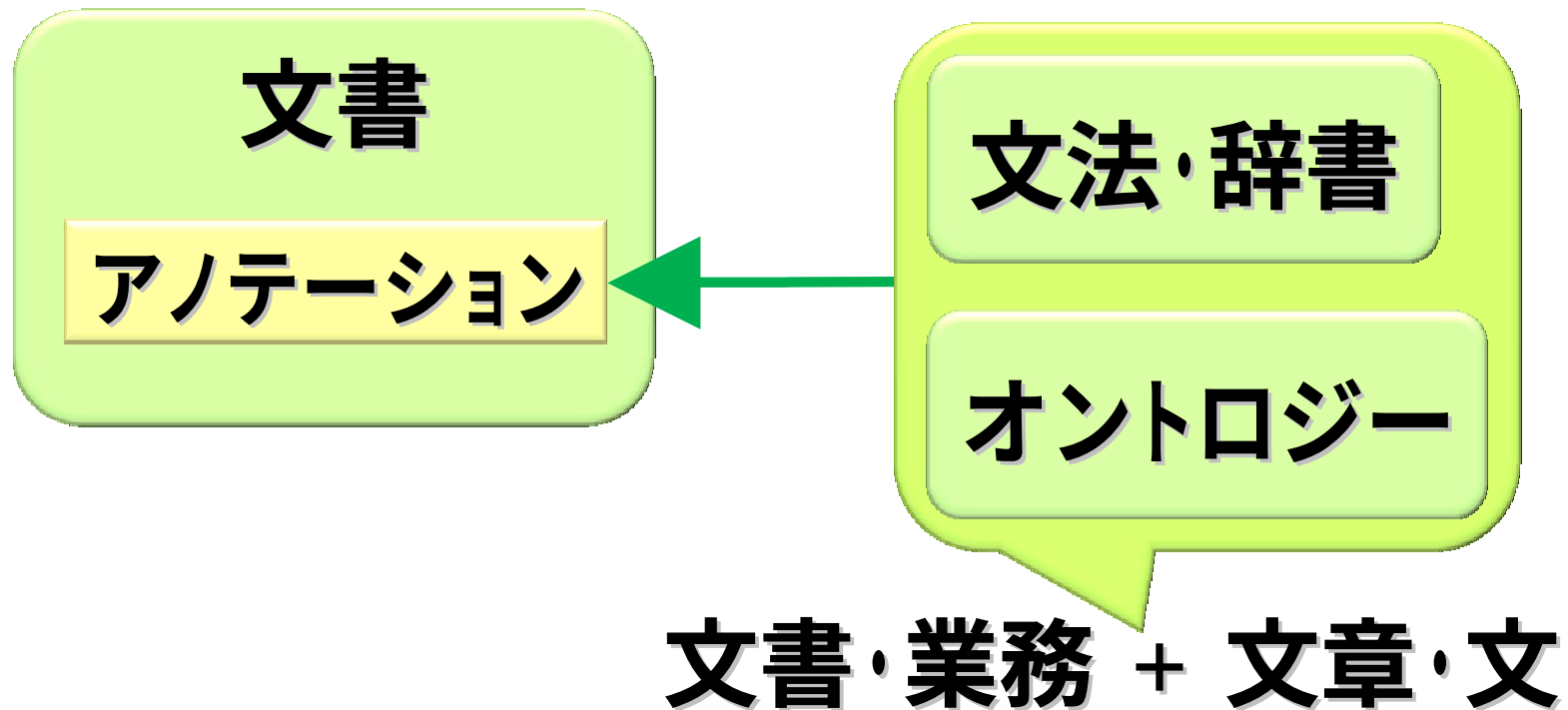
言語知識の体系化による 知的活動の支援

橋田 浩一

言語処理学会/産業技術総合研究所

産業日本語の設計

- 人間とコンピュータが意味を共有
 - 標準化によるデータ連携



平成22年度 JAPIO 特許版・産業日本語委員会

特許版・産業日本語委員会

年3回開催

文法・辞書

特許ライティングマニュアル 作成タスクフォース

機械翻訳に活用できる特許明細書ライティングマニュアル(案)の作成

年7回程度開催

タスクフォース長: 横井 (Japio研究所顧問)

知財、言語処理、知財翻訳、機械翻訳の現場実務者レベルでの共同作業

オントロジー

特許オントロジー 検討小委員会

利用

特許情報管理、検索等に役立つオントロジーの検討

年5回開催

小委員長: 橋田 (産総研)

知財、言語処理、検索の研究者で構成

特許ライティングマニュアル

- 明晰な日本語で特許文章を書くことに基
いて、効果的な特許文書を作成するための
マニュアル
- 大きく和文原稿ライティングと翻訳原稿ライ
ティングに分かれ、発明者・知財管理者・知
財翻訳者に利用される
- 特許ライティング支援システム（支援環境）
の漸進的な開発の第一歩
- 広く産業技術（文書・文章）ライティングのラ
イティングマニュアルの典型例
- 本マニュアルに基づいて何種類かの現場マ
ニュアルが制作される

特許ライティングマニュアルの構成

● 特許申請書類の品質向上

「文書」構造

- 「文書」document: コンテンツとしてのまとまり
- 技術文書としての分かり易さの観点
- 項目特性 (詳細な説明など) に応じた
記載内容、推奨表現



「特許文書ライティングマニュアル」

「文章」構造

- 「文章」text: 表現メディア (言語表現) としてのまとまり
- 明晰性の観点
 - 多言語への翻訳の観点
- 長文の分割、係り受けの改善
- 用語、用字の統一**
- 表記統一の観点



「特許文章ライティングマニュアル」

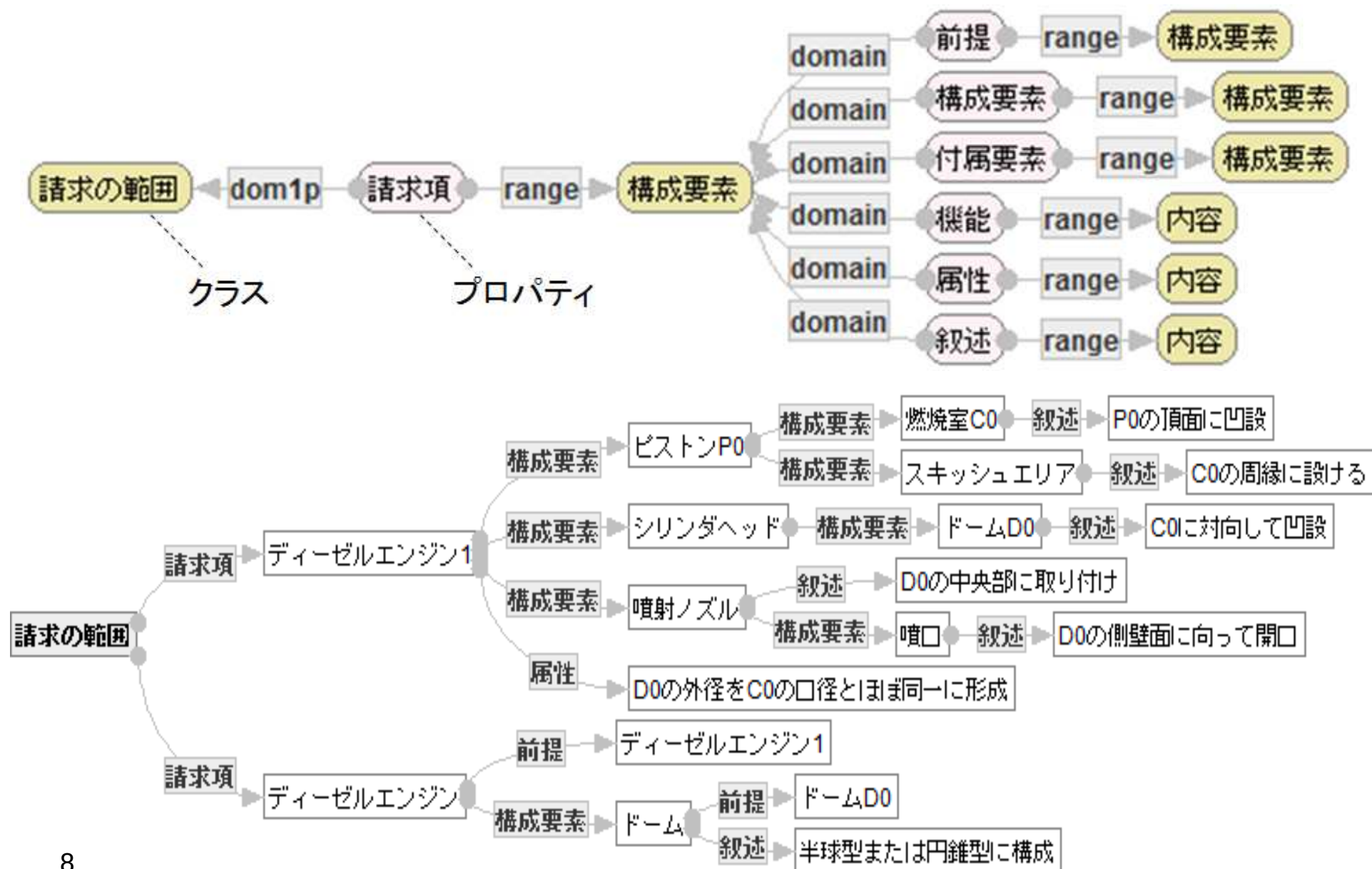
ライティングマニュアルに基づく言い換えによる翻訳の実験: データの蓄積

- 言い換え {180文, MT113文} /198文
- 言い換え規則 153個
 - 依頼原稿用 106個
 - 翻訳原稿用 47個
- 言い換え規則の分類 43
 - 依頼原稿用 19分類
 - 翻訳原稿用 24分類

作業のスナップショット

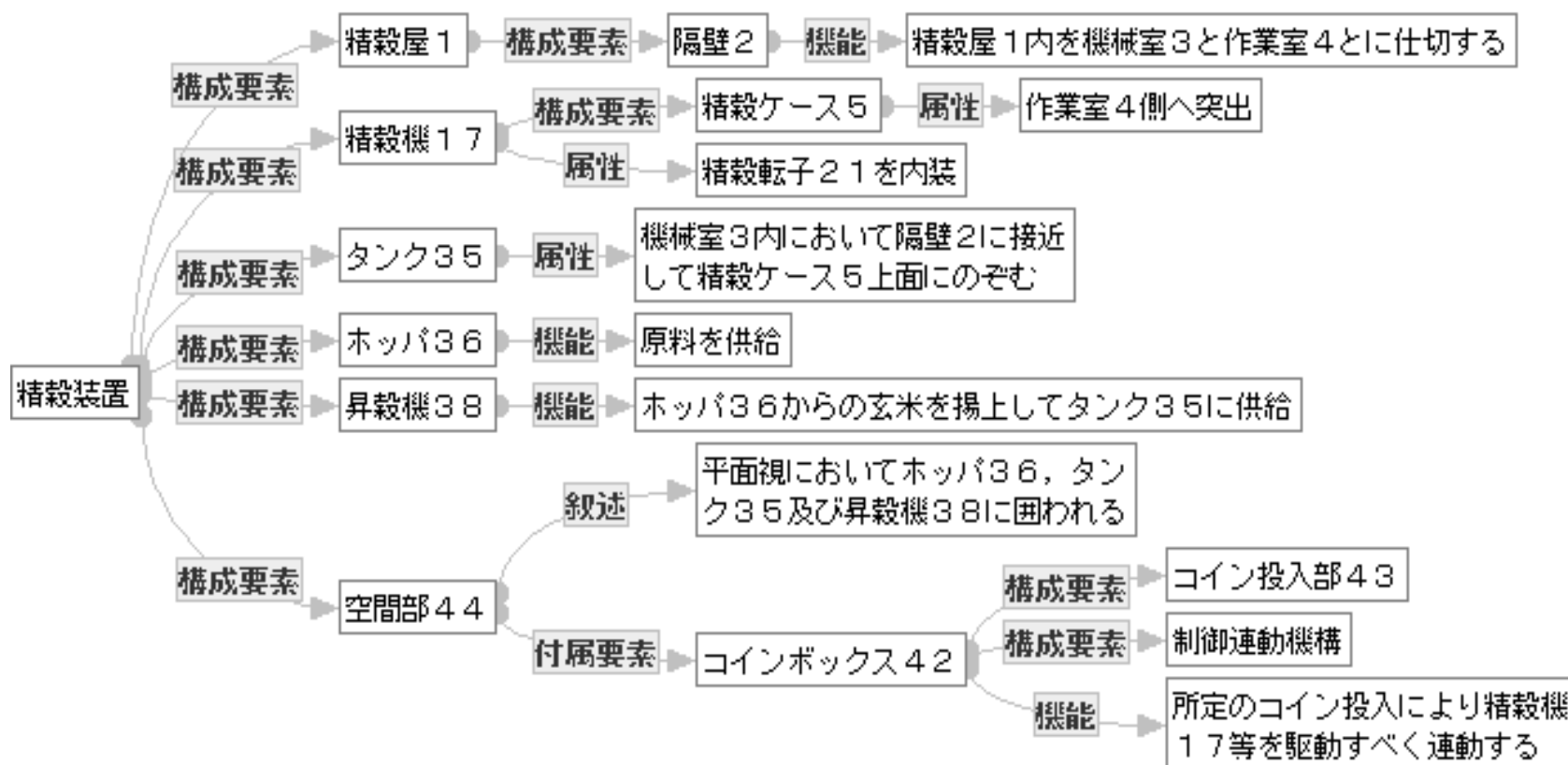
- [0022]-29 (原文から)
 - 本発明にあっては、インク容器の具体的な構造として、インク容器本体内の底部側にインク残液領域部を設け、このインク残液領域部の形状と容積が保形手段により保持される構造を採用したため、以下の作用効果が得られる。
- [言い換え]
 - 長文であるため、2文の連文に言い換え、接続詞を補う。すなわち、
「本発明にあっては、...構造を採用したため、以下の作用効果が得られる。」→「本発明にあっては、...構造を採用した。このことにより、以下の作用結果が得られる。」
 - 文語調表現を通常表現に言い換える。すなわち、
「本発明にあっては、」→「本発明は、」
 - 適切な時制に言い換える。すなわち、
「...採用した。」→「...採用している。」
 - 並立表現を整えるように言い換える。すなわち、
「形状と容積が」→「形状と容積とが」
 - 語調を整える表現を削除し、冗長表現を適正な表現に言い換える。すなわち、
「保形手段により保持される」→「保形手段に保持される」
- [依頼原稿和文]
 - 本発明は、インク容器の具体的な構造として、インク容器本体内の底部側にインク残液領域部を設け、このインク残液領域部の形状と容積とが保形手段に保持される構造を採用している。このことにより、以下の作用効果が得られる。

特許オントロジーとそのインスタンス



書き流し型請求項の例

精穀屋1内を機械室3と作業室4とに仕切る隔壁2を通して、内部に精穀転子21を内装する精穀機17の精穀ケース5をこの作業室4側へ突出し、上記機械室3内において隔壁2に接近して精穀ケース5上面にのぞむタンク35及び原料を供給すべきホッパ36とを配設すると共に、該ホッパ36からの玄米を揚上してタンク35に供給すべく昇穀機38を設け、平面視においてこれらホッパ36、タンク35及び昇穀機38に囲われる空間部44には、コイン投入部43と制御連動機構とを有して所定のコイン投入により精穀機17等を駆動すべく連動するコインボックス42を配設してなる精穀装置。



コンテンツのライフサイクルと意味構造化

- 作成者にとって自明な構造を作成者が明示するインセンティブは小さい。
 - 特許の請求項、説明書、論文、etc.
- しかし、多くの利用者にとってその構造がわかりにくいため利用価値が低い。
 - 読解、翻訳、検索、etc.
 - インフラ ... 明確に説明できないと売れない
- コンテンツの価値を最大化するように業務を改革することが重要。
 - ステークホルダの間での価値の共有
 - 技術による著者の教育(行動変容)

データ生成時の構造化: 診療録の選択入力

入院患者管理		病名一覧	
トクガワ イエヤス	2000年12月12日		
徳川 家康	男 111-01-00001		
コピー トクガワ イエヤス 徳川 家康 平成12年12月12日 生 111-01-00001 1 2 3 - 1 2 3 4 東京都大田区蒲田 1 - 2 - 3 0 3 - 1 2 3 4 - 1 2 3 4		整 <感染症、寄生虫症> <新生物> <血液、造血器疾患、免疫機構障害> <内分泌系疾患、代謝疾患> <精神、行動障害> <神経系疾患> <眼、付属器疾患> <耳、乳突炎疾患> <嗅覚器系疾患> <呼吸器系疾患> <消化器系疾患> <皮膚、皮下組織疾患> <泌尿生殖器系疾患> <運動器系疾患> <妊娠分娩、産褥> <先天奇形変形、染色体異常> <症状徴候、異常臨床、所見非分類> <損傷中毒、その他外因影響>	
病名入力 入院時病名		☐ 右変形性股関節症 ☐ 急性胃炎	
主訴 歩行障害、右膝関節痛		☒ 右変形性股関節症	
現症 ☆セル内での改行は41 平成20年1月1日より 今回は右TKRの目的 ☆セルを変更、修正し		☐ 急性胃炎	
既往歴 08/05/14 急性胃炎		☐ 慢性腎臓病	
現病歴 09/03/14 胃痛とめま 09/05/01 111-01-000 上記画像が 08/05/05 胸部痛があった。また、血圧も高かった。 111-01-00001wri2009050501 特に異常は認められない。		☐ 慢性腎臓病 ☐ 急性腎炎 ☐ 慢性糸状腎炎 ☐ ネフローゼ症候群 ☐ 腎臓不全 ☐ 腎臓結石症 ☐ 尿管結石症 ☐ 過剰腎 ☐ 糖尿病 ☐ 尿毒症 ☐ 腎性貧血 ☐ 尿路感染症 ☐ 前立腺肥大症 ☐ 前立腺炎 ☐ 亀頭包皮炎 ☐ 乳腺炎 ☐ 女性化乳房症 ☐ 乳腺腫瘍 ☐ 月経困難症 ☐ 更年期症候群 ☐ 妊娠分娩、産褥 ☐ 子宮外妊娠 ☐ 悪阻 ☐ 閉経リウマチ ☐ 関節炎 ☐ 総動脈硬化 ☐ 痛風 ☐ 左変形性股関節症 ☐ 右変形性股関節症 ☐ 右変形性股関節症 ☐ 関節痛 ☐ 骨内腫 ☐ 膠原病 ☐ シェーグレン症候群 ☐ 全身性エリテマトーデ ☐ 骨髄嚢腫 ☐ 変形性骨粗鬆症 ☐ 頸椎症候群 ☐ 坐骨神経痛 ☐ 椎間板ヘルニア ☐ 腰痛症 ☐ がんグリオン ☐ 結核菌 ☐ 肺動脈周囲炎 ☐ 脳炎 ☐ 脊髄上神経痛 ☐ 後頭神経痛 ☐ 上肢神経痛 ☐ 神経痛 ☐ 背腰部筋痛 ☐ 骨粗鬆症 ☐ 胸郭変形 ☐ 骨癒死症 ☐ 尿酸性腎臓病 ☐ 急性腎炎 ☐ 慢性糸状腎炎 ☐ ネフローゼ症候群 ☐ 腎臓不全 ☐ 腎臓結石症 ☐ 尿管結石症 ☐ 過剰腎 ☐ 糖尿病 ☐ 尿毒症 ☐ 腎性貧血 ☐ 尿路感染症 ☐ 前立腺肥大症 ☐ 前立腺炎 ☐ 亀頭包皮炎 ☐ 乳腺炎 ☐ 女性化乳房症 ☐ 乳腺腫瘍 ☐ 月経困難症 ☐ 更年期症候群 ☐ 妊娠分娩、産褥 ☐ 子宮外妊娠 ☐ 悪阻	

自由入力 of 構造化 → 2次利用

- 病理診断報告書の入力支援 → 検索・分析

入力テキスト

(laparoscopic partial resection.)

4.2x1.5x1.1cm

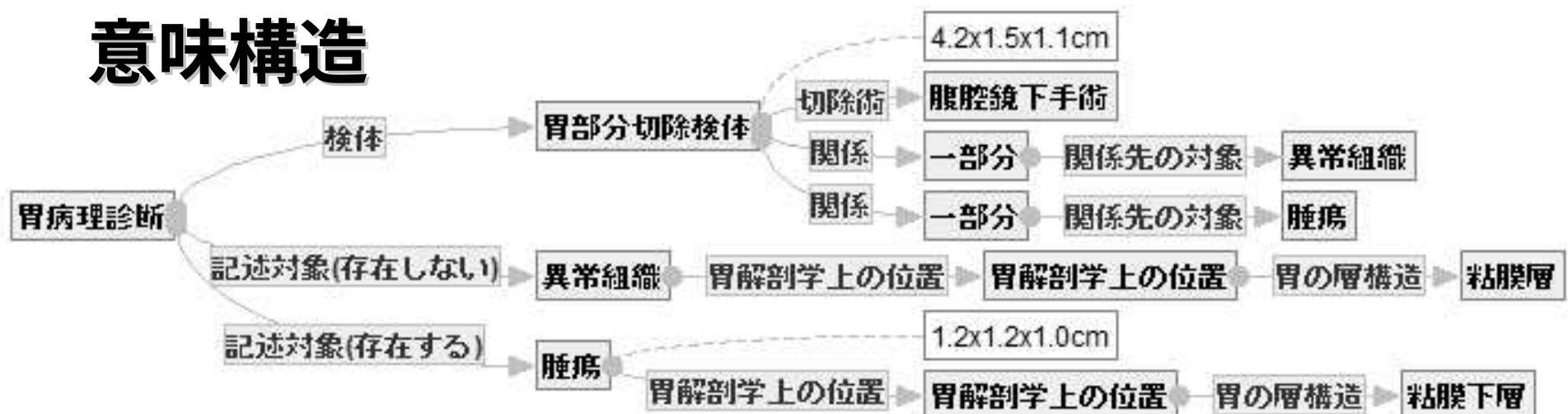
粘膜面に著変なし. 粘膜下層に主座をおく
粘膜下腫瘍1.2x1.2x1.0cm大あり.

主座をおく **ねん**

粘膜下腫瘍

入力補完

意味構造



オントロジーのさまざまな利用

- オントロジーに基づくグラフ構造を用いた検索、翻訳、要約など
- 発明概念の抽出
 - 請求項と実施例・効果との対応関係
 - 発明概念を中心とする読解支援

特許の分類体系

- IPC
- FI
- Fターム
- 巨大な体系だが使い切れていない?
 - 横断的な知の体系としての利用法が可能かも
- Fタームとオントロジーとの融合
 - 観点は属性(property)
 - オープンオントロジーの中の構造化と関連するか

Fターム

- テーマグループ:41個 例) 2B
- テーマ:約3000個
 - 例) テーマコード:2B002 テーマ名:化粧合板 FI
カバー範囲:B27D5/00
 - <テーマ> <is-instance-of> <テーマグループ>
 - <テーマ> <FI適用範囲> {FI記号}
- 観点:テーマごとに定義
 - 例) テーマ「人工魚礁」における観点 AA00「設置形態」
 - <観点> <FI適用範囲> {FI記号}
- Fターム:テーマごとに定義
 - <Fターム> <観点> <テーマ>
 - 例) テーマ「人工魚礁」における観点AA00「設置形態」のタームA01「沈設型」
 - <Fターム> isa <Fターム> /ドット階層

Fタームの例

2B003

人工魚礁 自然資源

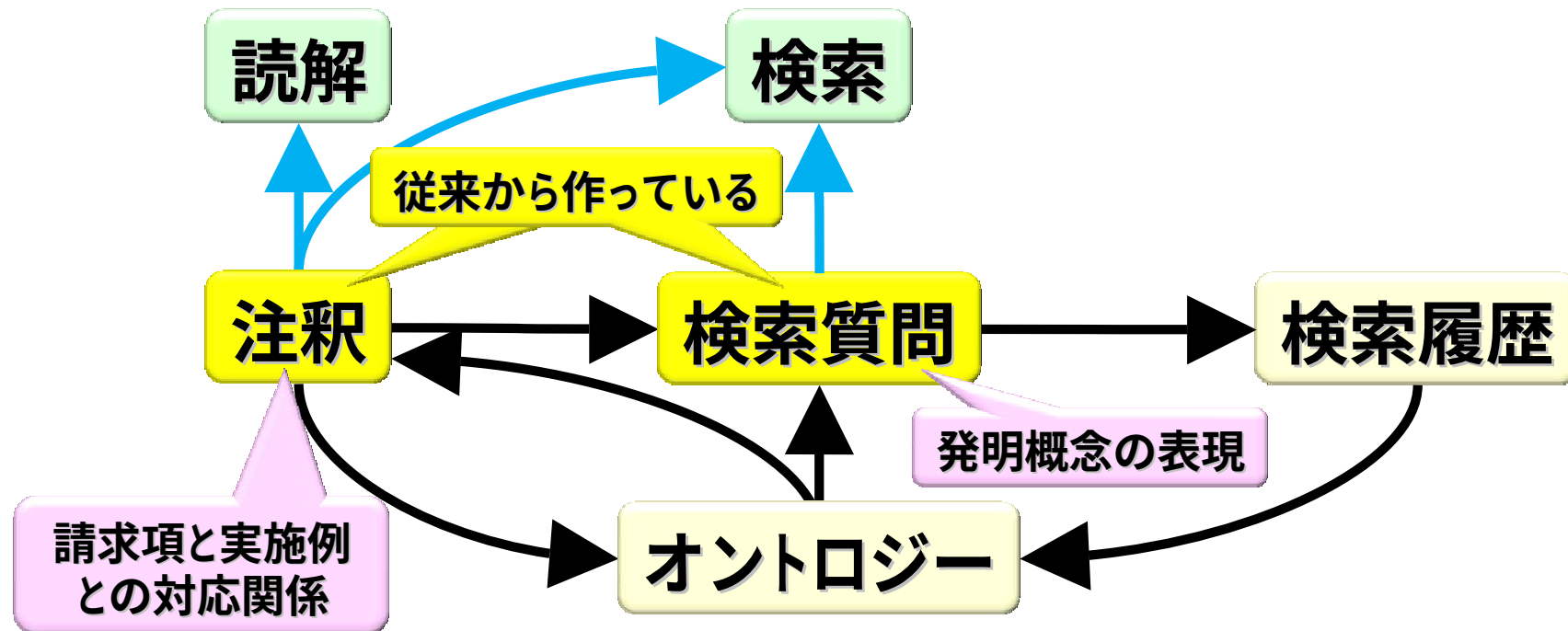
A01K61/00,311-61/00,321

観点		Fターム									FI適用範囲	
AA	AA00										<u>A01K61/00,311-61/00,321</u>	
	設置形態	・沈設型	・浮沈型	・浮き型	・護岸							
BB	BB00	BB01	BB02	BB03	BB04	BB05	BB06	BB07	BB08	BB09	<u>61/00,321</u>	
	構造	・ブロック (一体成型のみ)	・組立式	・ブロック	・棒状体	・筒状体	・板状体	・板体に穴を形成したもの	・連結部に特徴があるもの	・人工海藻		
CC	CC00	CC01	CC02	CC03	CC04	CC05	CC06					<u>1</u>
	構造以外の 特徴	・湧昇流発生	・表面処理	・誘引手段 (含:肥料)	・設置方法 (含:装置)	・製造方法	・起上り小法師					
DD	DD00	DD01	DD02	DD03	DD04	DD05	DD06	DD07	DD08	DD09		
	材質	・コンクリート	・金属	・合成樹脂	・産業廃棄物	・タイヤ	・石、砂	・陶磁器	・貝殻	・磁石		
EE	EE00	EE01	EE02	EE03	EE04							
	対象生物	・アワビ (含:貝)	・イセエビ (含:エビ)	・イカ、タコ、ウニ	・海藻							

<http://www5.ipdl.inpit.go.jp/pmgs1/pmgs1/!frame?hs=1&gb=2&dep=3&sec=2B&cls=003&scls=&mgrp=&idx=&sgrp=&sf=&bs=&dt=&wrd=&nm=>

集合知に基づく先行技術調査 のアーキテクチャ

- 特許文書への注釈と検索質問と辞書を通じて調査者の知識が循環



ソーシャルeサイエンス

生活、産業、行政、
医療、etc.

社会と融合したeサイエンス

- 社会が定常的に生成するデータや知識を用い、社会の営みと融合した大小の仮説検証サイクル
- データの意味的構造化と相互連携によって1次利用の利便性と2次利用(検索・分析など)における有用性を同時に高める

