

# セマンティックコンピューティング

産業技術総合研究所

情報技術研究部門

橋田 浩一

2006-02-07

## セマンティックギャップ

- 人間とコンピュータが意味を共有していない。
  - コンピュータが人間を理解できない。
  - 人間がコンピュータを理解できない。
- うまく協調できない。

# コンピュータが人間を理解できない

- 検索エンジンで欲しい情報が見付からない。無関係な情報が大量に出てきて関係のある情報がほとんど出てこない。
- 人間が何を知りたいかがコンピュータに正しく伝わっていない。
- Webサイトを使いやすく構成し維持するのが大変だ。
- Webコンテンツの意味がコンピュータにわからない。
- 社内メールを禁止したら業績が上がった。
- コミュニケーションの背景や文脈を電子メールが十分に反映していない。

3

# 人間がコンピュータを理解できない (コンピュータも自分を理解できない)

- Windows XPにサービスパック2を入れたら、どういうわけか無線LANにつながらなくなった。
- MS Wordを使っていたら妙な点線が出てきて消せない。
- イン트라ネットの中で業務ワークフローの間の連携が取れない。
- デジタル編集技術が普及しない。

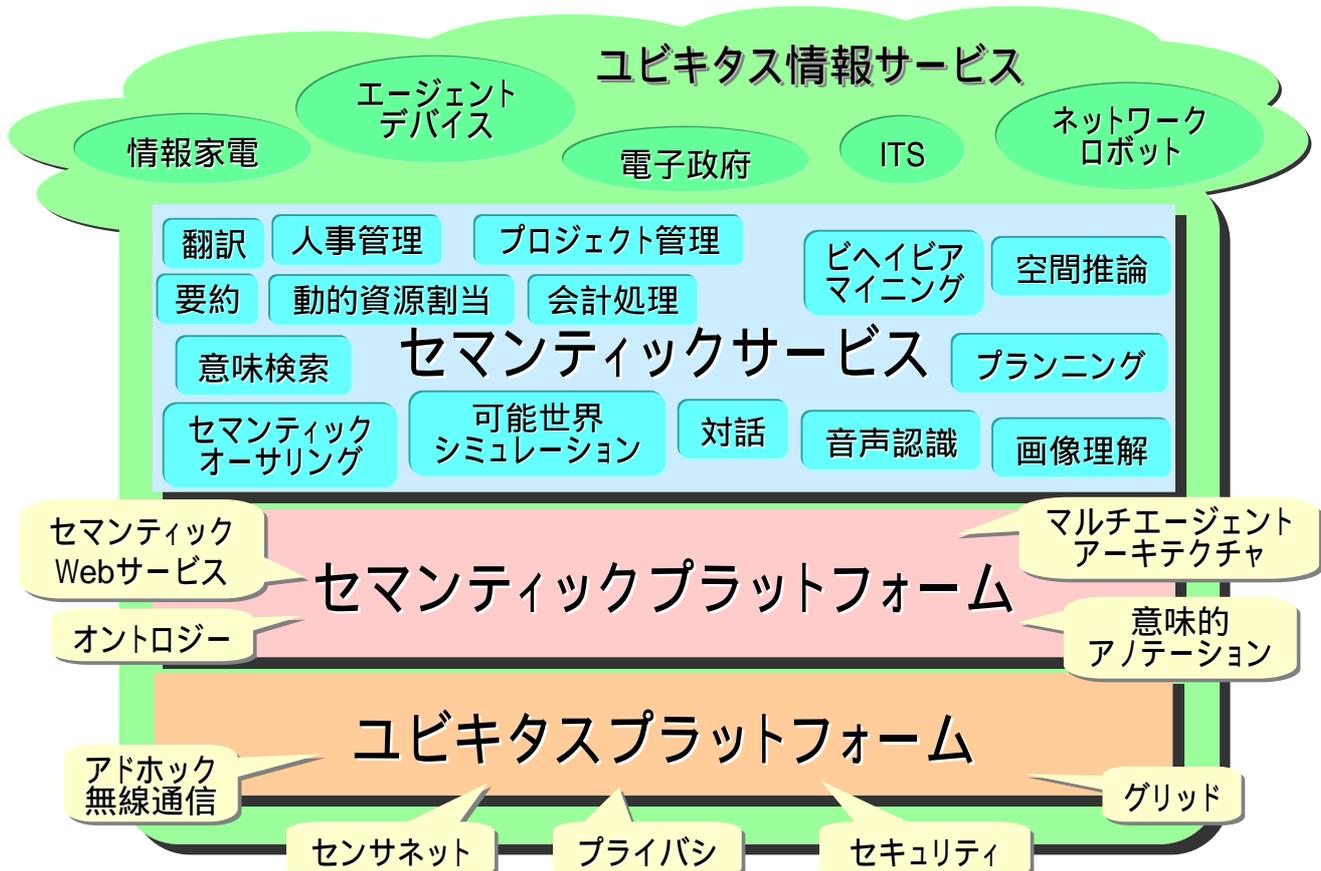
4

# セマンティックコンピューティング

## = Semantics-Oriented Architecture

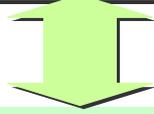
- ガラス張りコンピュータ
  - 人間に理解できる意味に基づいてコンピュータシステムを設計・運用
  - データモデル + プロセスモデル
- 利用者に有意味なサービスを提供
  - 設計の最初から有意味
  - 検索要求を検索エンジンに正確に伝えられる
- 利用者がシステムを理解・作成・改良
  - 利用者による改善 創発的最適化

5



6

セマンティックギャップ



対話グラウンディング

シンボルグラウンディング

ITシステムの扱う  
意味を人間が理解

人間の扱う意味を  
ITシステムが理解

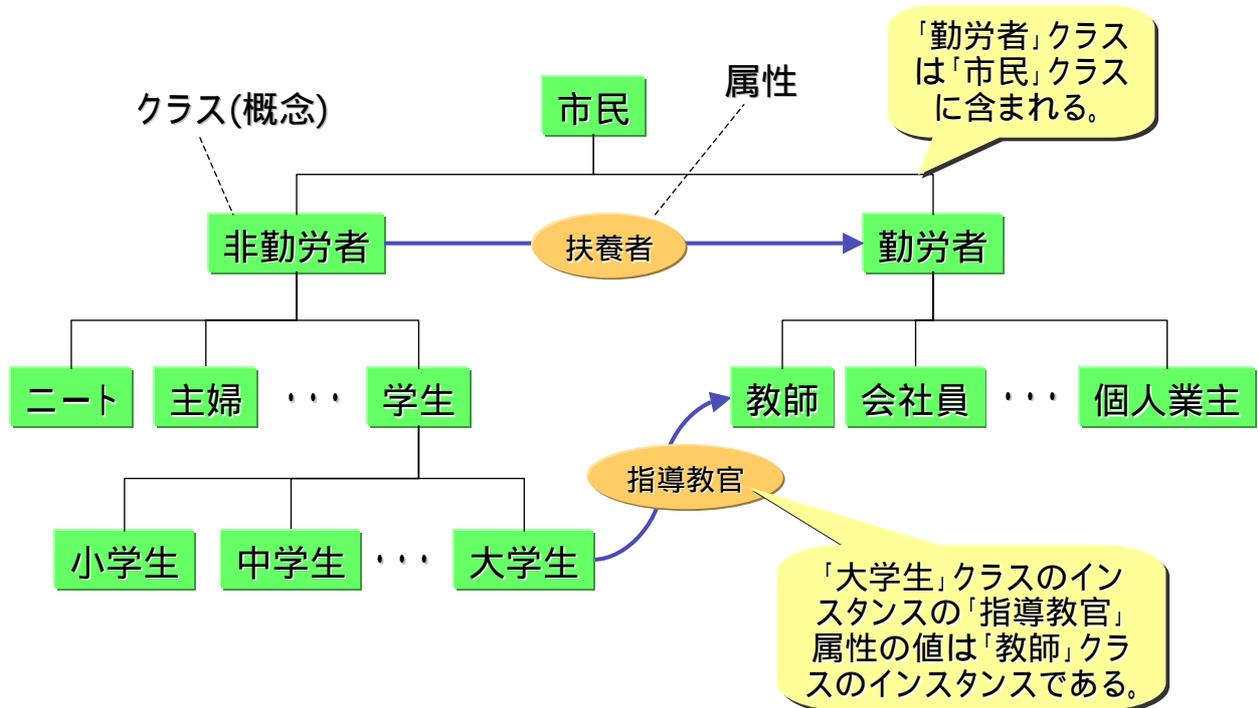
実世界の  
意味の理解

common groundの確立

7

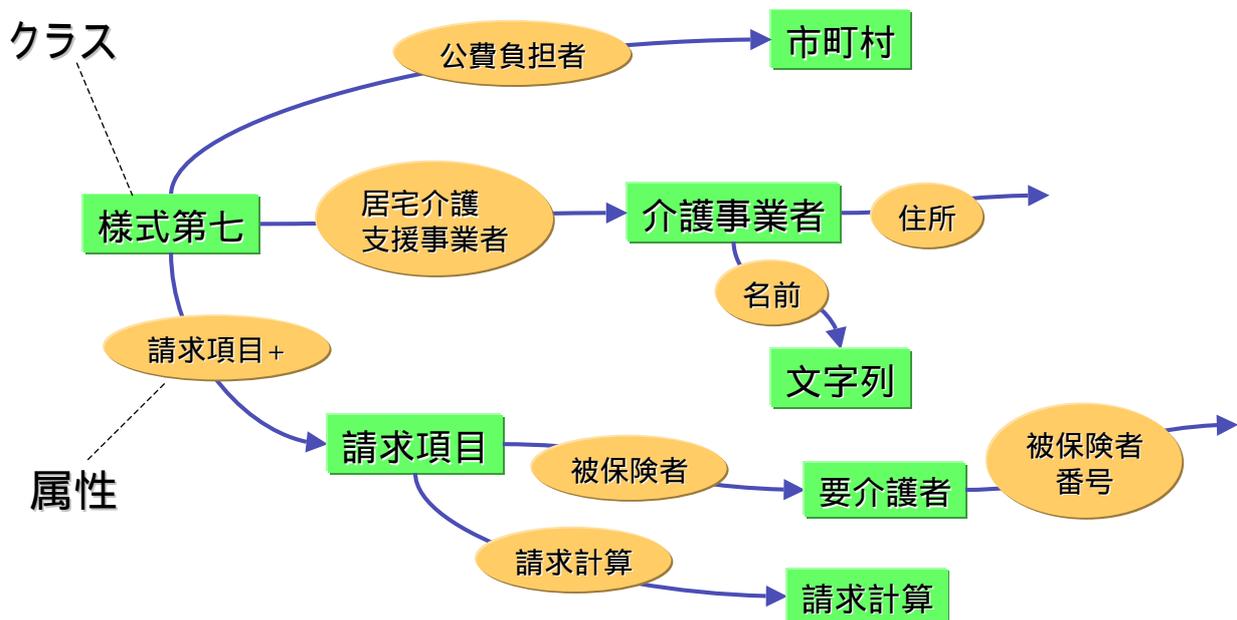
定型的な意味構造

# オントロジー = 概念階層 + 属性定義



9

# 文書のオントロジー



10

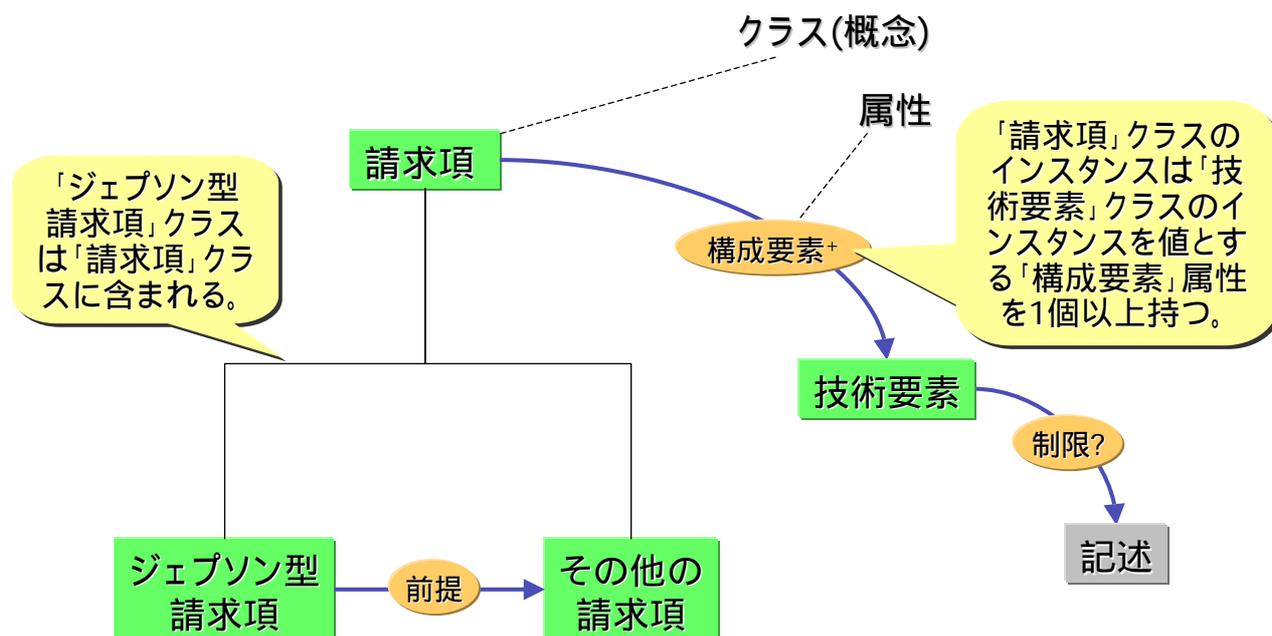
居宅介護支援介護給付費明細書

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					〒					
	事業所名称					連絡先					電話番号					
単位数単価										▲		(円/単位)				

項番	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														
	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	から	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														
	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	から	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														
	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	から	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								

非定型的/詳細な意味構造

# 特許の請求項のオントロジー



13

## 請求項の例

イオン源(1)と、このイオン源(1)から引き出されたイオンを質量分離する質量分析部(2)と、この質量分析部(2)で質量分離して取り出されたイオンを電子に変換するイオン電子コンバータ(4)と、変換された電子を検出して電気信号として取り出す電子検出器(3)とを備える質量分析装置において、前記質量分析部(2)とイオン電子コンバータ(4)との間に配置されたサブスリット(10)と、イオン源(1)に印加するイオン引出電圧を $V_0$ 、サブスリット(10)に印加する電圧を $V_s$ 、イオン電子コンバータ(4)に印加するコンバータ電圧を $V_c$ とした場合に、 $V_s = V_0 - k_1$ 、 $V_c = V_0 - k_2$  (ただし、 $k_1$ 、 $k_2$ は定数)を満たすように、 $V_0$ の設定に応じて $V_s$ 、 $V_c$ をそれぞれ決定する電圧制御手段(12)を備えることを特徴とする質量分析装置

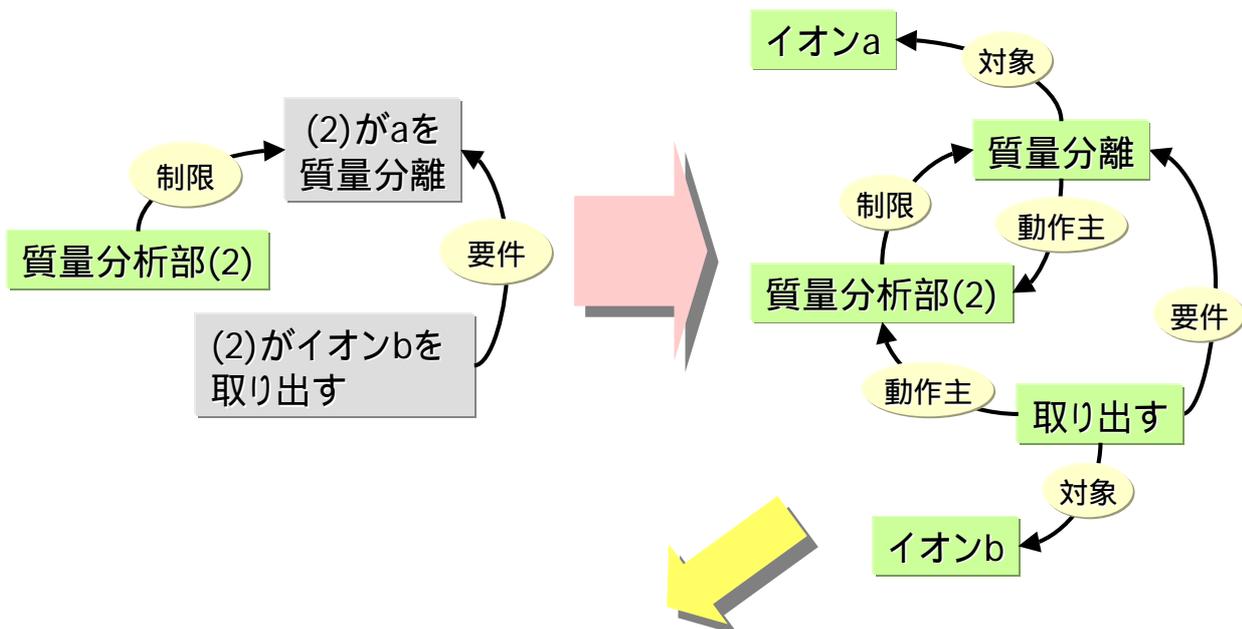
14

# オントロジーのインスタンスとして記述された請求項の内容



15

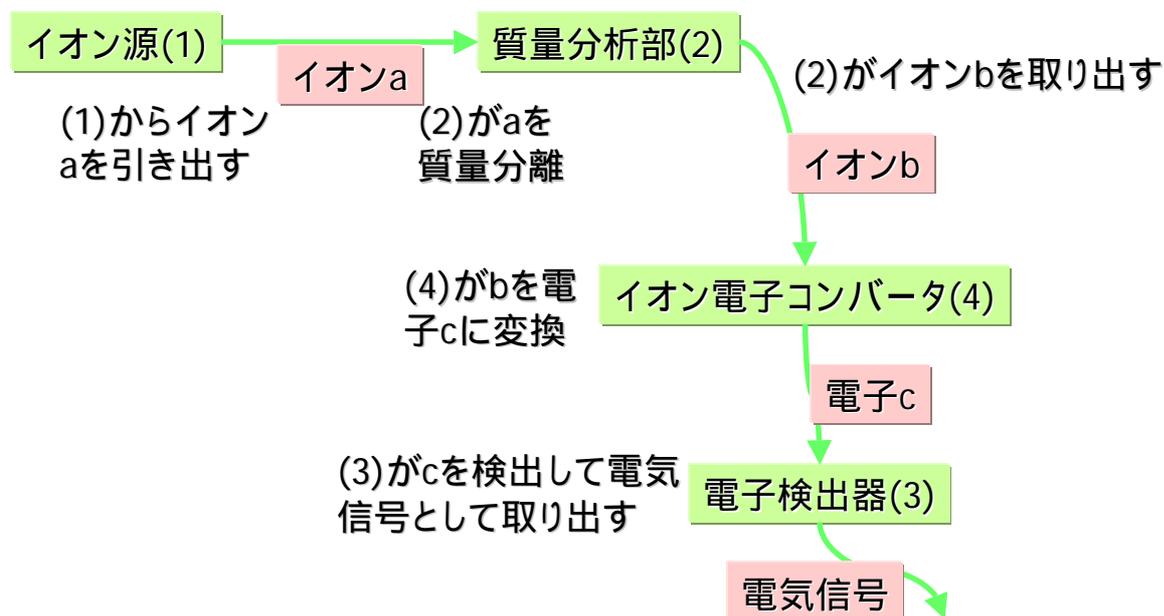
# 自然言語処理による解析・詳細化



高精度の検索、要約、言い換え、翻訳

16

# フローチャートの自動生成



17

## 翻訳とそのチェック・・・2日間

検索質問Qのノードxごとに、リンクy-zがデータベースDに含まれてyのラベルがLであるようなノードyとノードz  $F(x)$ が存在するような、ラベルLのリストを、表示部に表示する

### 誤った翻訳

displaying, on a display unit, a list of labels L in which are present a node z  $F(x)$  and a node y of which a link y-z is contained in the database D and of which the label y is L, for each of the nodes x of a search question Q

18

# 意味構造による内容の明瞭化 2時間でチェック可能

検索質問Qの各ノードx

量化

Lのリストを表示部に表示する

内包

$z = F(x)$ .  
データベースDがリンクy-zを含む。  
yのラベルがLである。

each node x in  
retrieval query Q

quantify

to display the list of L on the display unit

intension

$z = F(x)$ .  
Database D contains link y-z.  
The label of y is L.

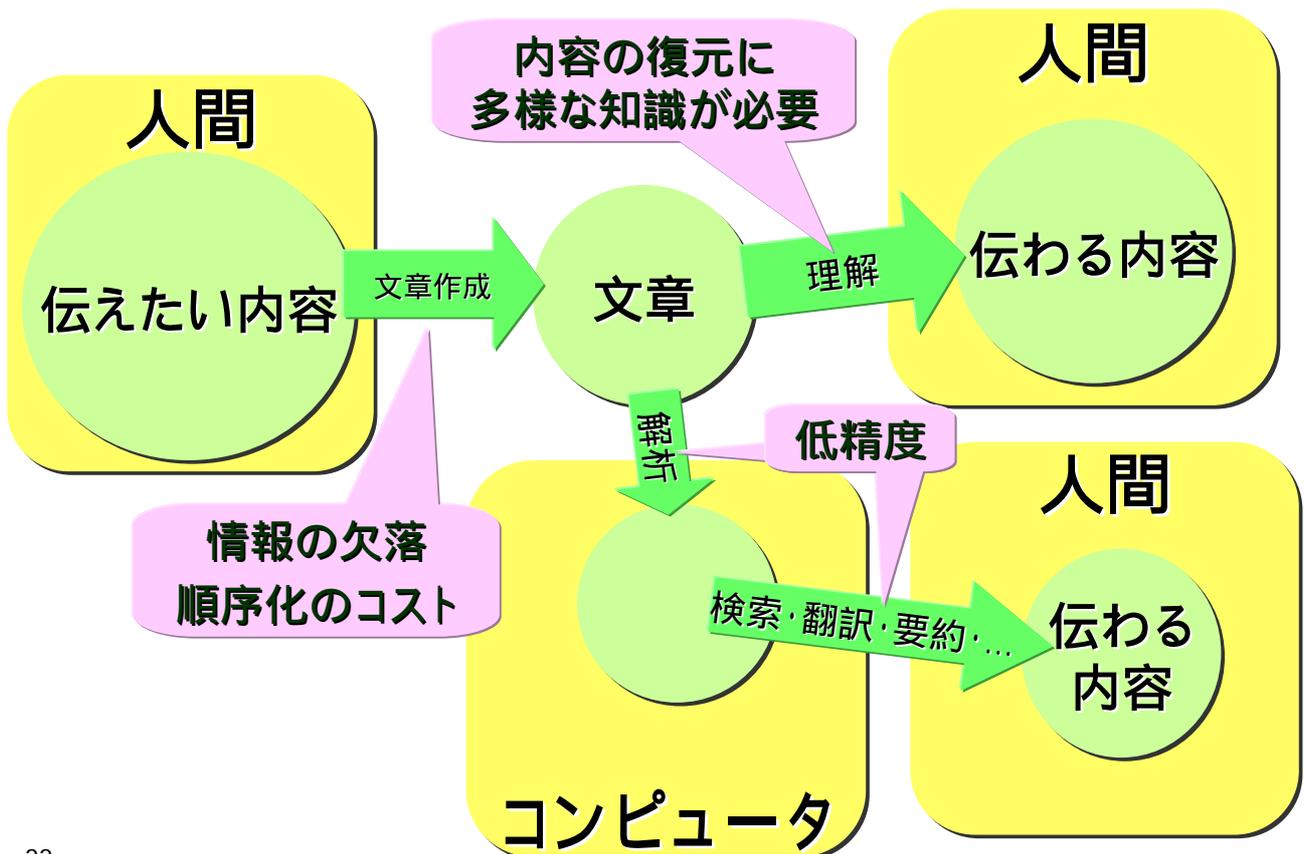
## セマンティックオーサリング

## 課題

- ~~いかにして多くの利用者に(機械の代わりに)意味構造化作業をさせるか?~~
- いかにして多くの利用者の知的生産性を向上させるか?
- 人間を創造に集中させる。

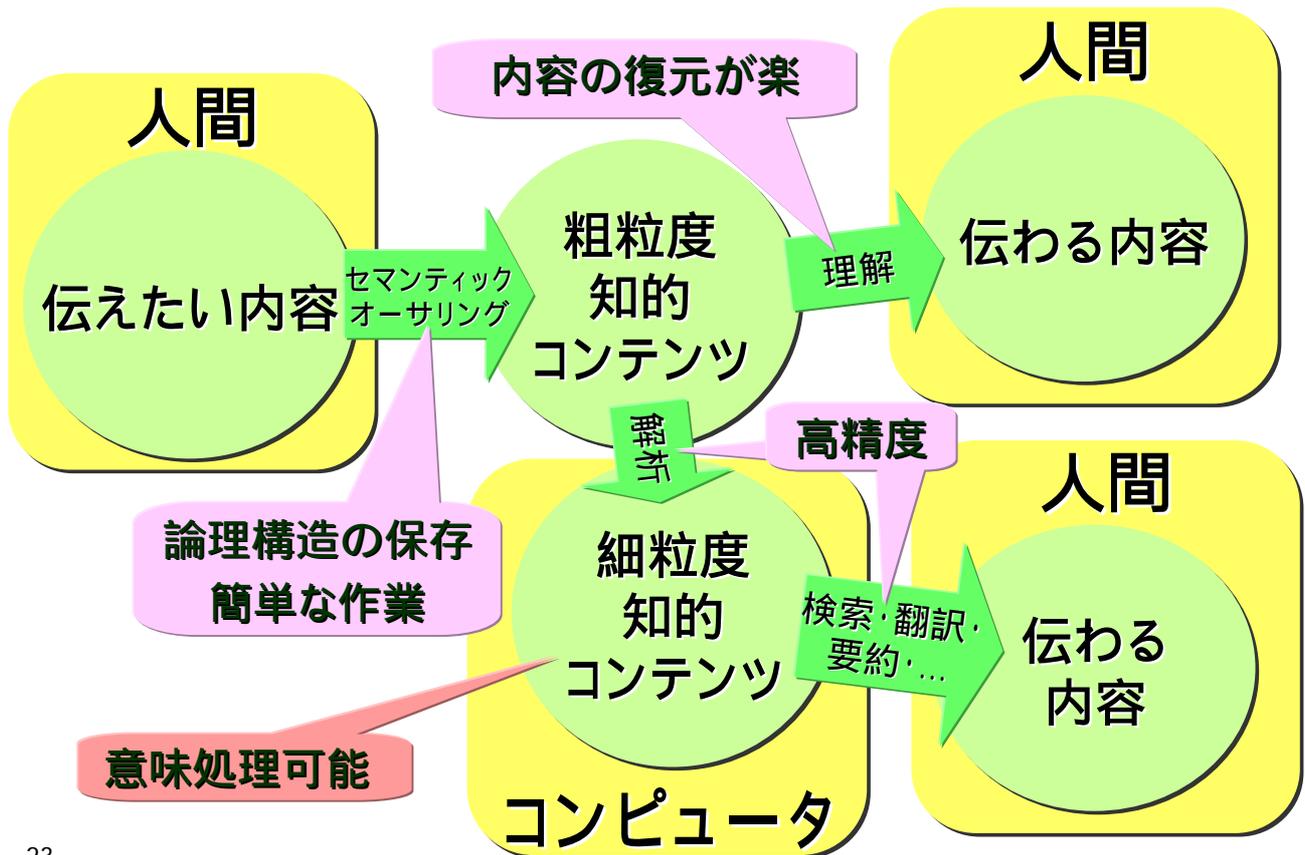
21

## 従来の文章による情報伝達



22

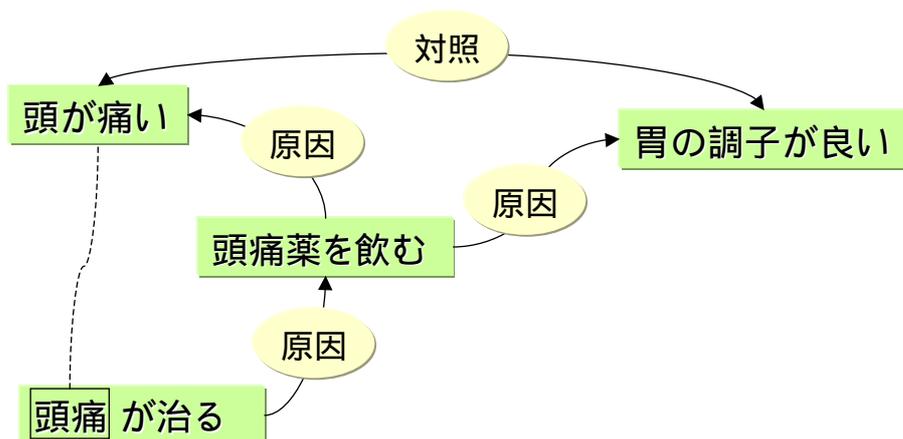
# 知的コンテンツによる情報伝達



23

## 粗粒度知的コンテンツ

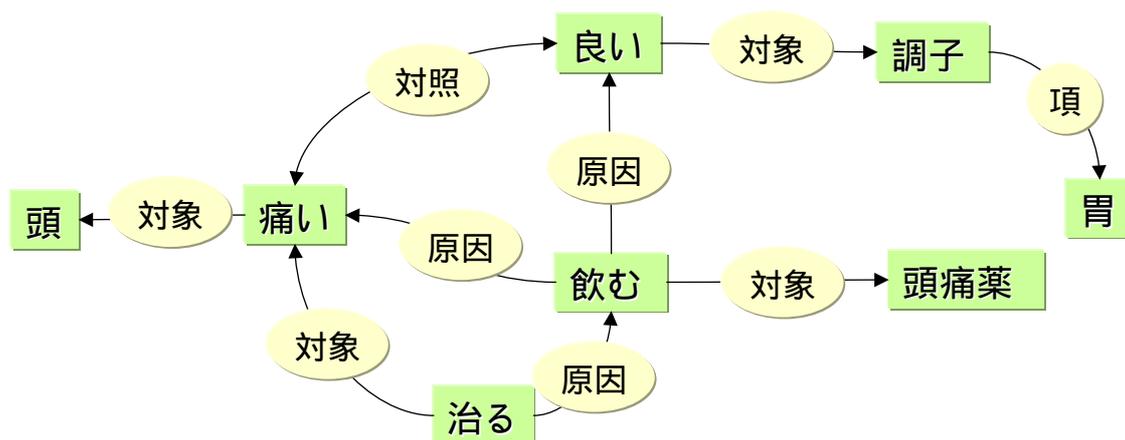
- セマンティックオーサリングの結果
- 人間がわかりやすく作りやすい・・・発想支援
  - 伝えたい内容の構造を保存
  - 文の順序等を決めなくてよい



24

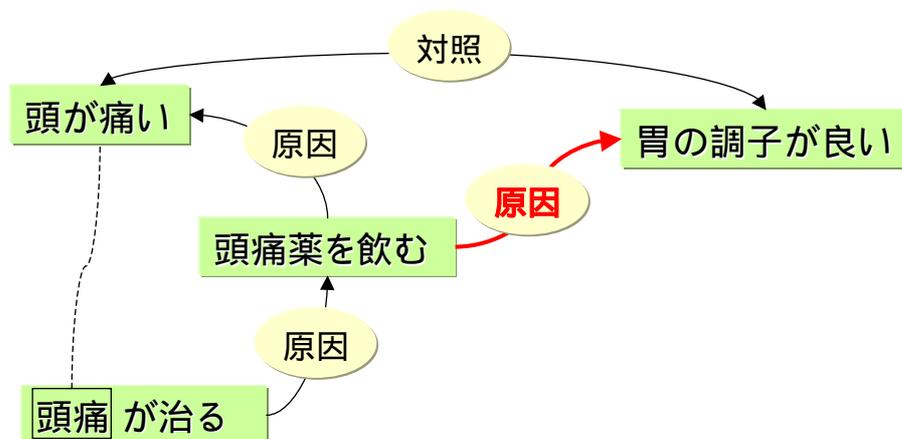
## 細粒度知的コンテンツ

- 粗粒度知的コンテンツの自動的詳細化
- 検索、要約、翻訳など
- 対話グラウンディングには粒度が過小



25

## セマンティックオーサリングは文章作成より簡単(1/2)



26

# セマンティックオーサリングは文章作成より簡単(2/2)

- 前頁の知的コンテンツと同じ内容の文章:

胃の調子良かった、でも頭が痛かった  
ので頭痛薬を飲んだら治った。

- 前頁の**原因**に相当する関係\*を反映するように書き換えるのは面倒:

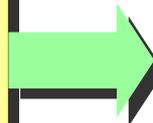
胃の調子良かったが頭が痛かった。  
そこで頭痛薬を飲んだら頭痛が治った。

27

## 発想支援システムによる文章の品質向上

### 八木下 他 (1998)の実験

発想支援システム  
でネットワーク型コ  
ンテンツを作成



ネットワーク型コ  
ンテンツに基づ  
いて文章を作成

- 見落としが少ない
  - 含まれる論点が多い
- 考えが深まる
  - 推論の連鎖が長い

28

## 従来の発想支援システム

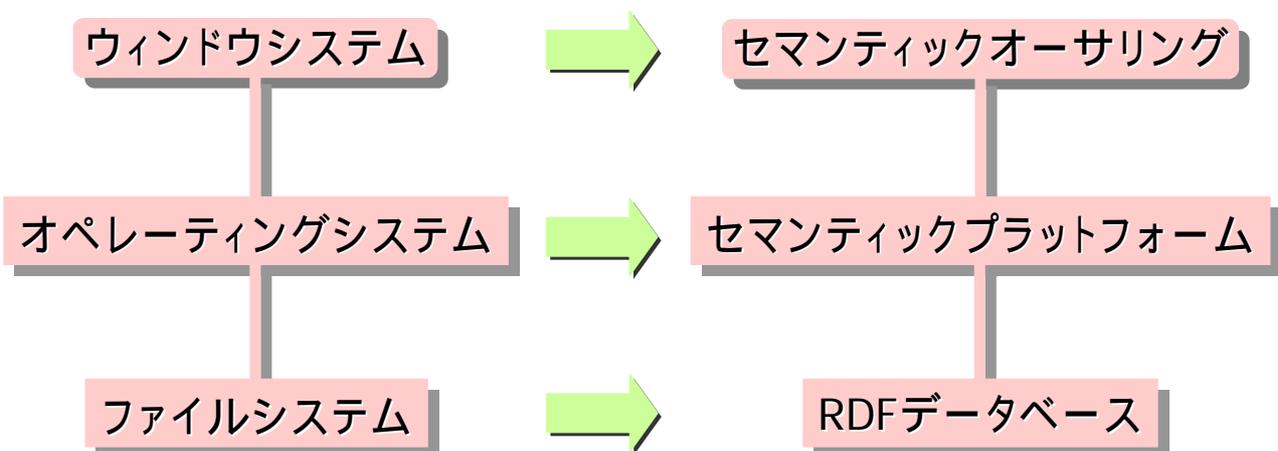
- 意味的關係が規格化されていない
  - プレスト参加者か著者本人しかわからない
  - 共有・再利用が困難
- 文章作成の手間
  - 見かけ上のコストが大きい 普及せず

## セマンティックオーサリング

- 意味的關係の規格化
  - 共有・再利用が容易
  - 高精度の検索、要約、翻訳など
- 文章生成の自動化
  - コストが小さい 普及

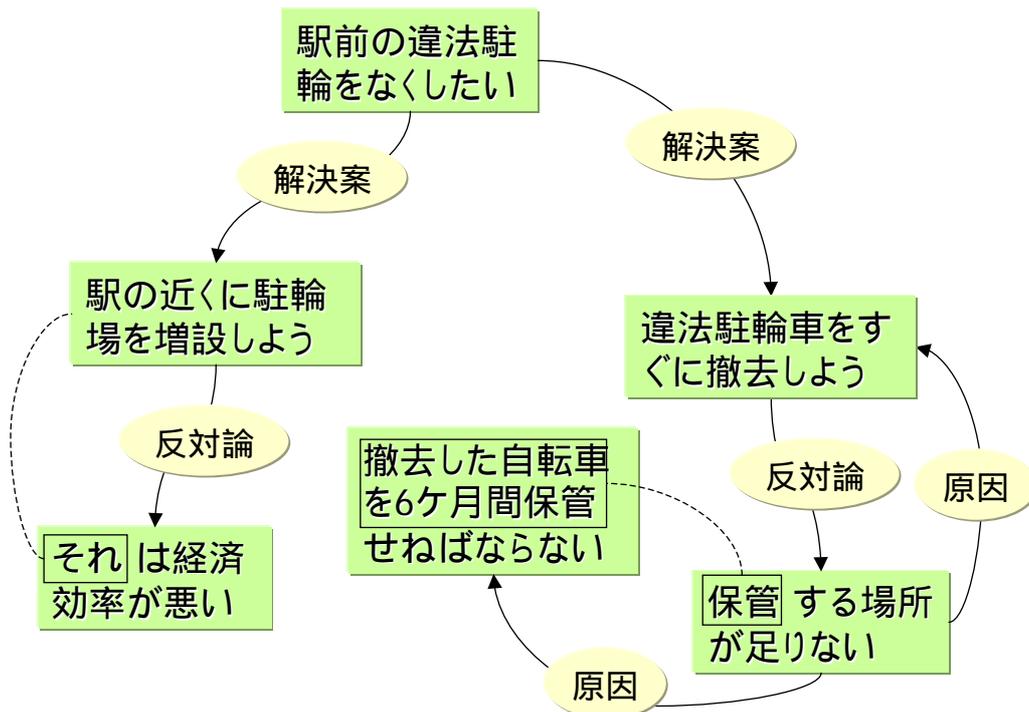
29

## セマンティックプラットフォームのリッチクライアントとしてのセマンティックオーサリング



30

# 共同セマンティックオーサリング



## 従来のグループウェア

- グループワークにしか使えない  
普及せず

## 共同セマンティックオーサリング

- 日常業務(個人用セマンティックオーサリング)とグループウェアのシームレスな融合
- グループウェアのメリット  
+ 高度な検索、要約、翻訳など

33

## デスクワークの生産性向上

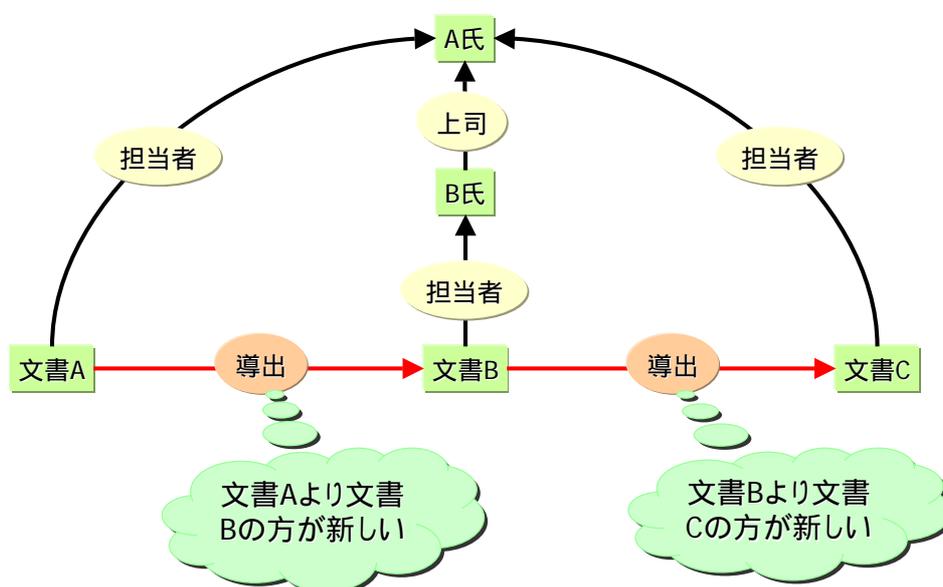
- 検索
  - 名寄せの共有 検索の多くは単なるリンク辿り
- 対応付け
  - リンクの管理と共有
- 読解
  - 明示的な論理構造 根拠の提示など
- 発想
  - 読解支援、検索、要約、可視化、etc.

例: 多数の関連する電子メールを収集して最新の状況を把握する作業

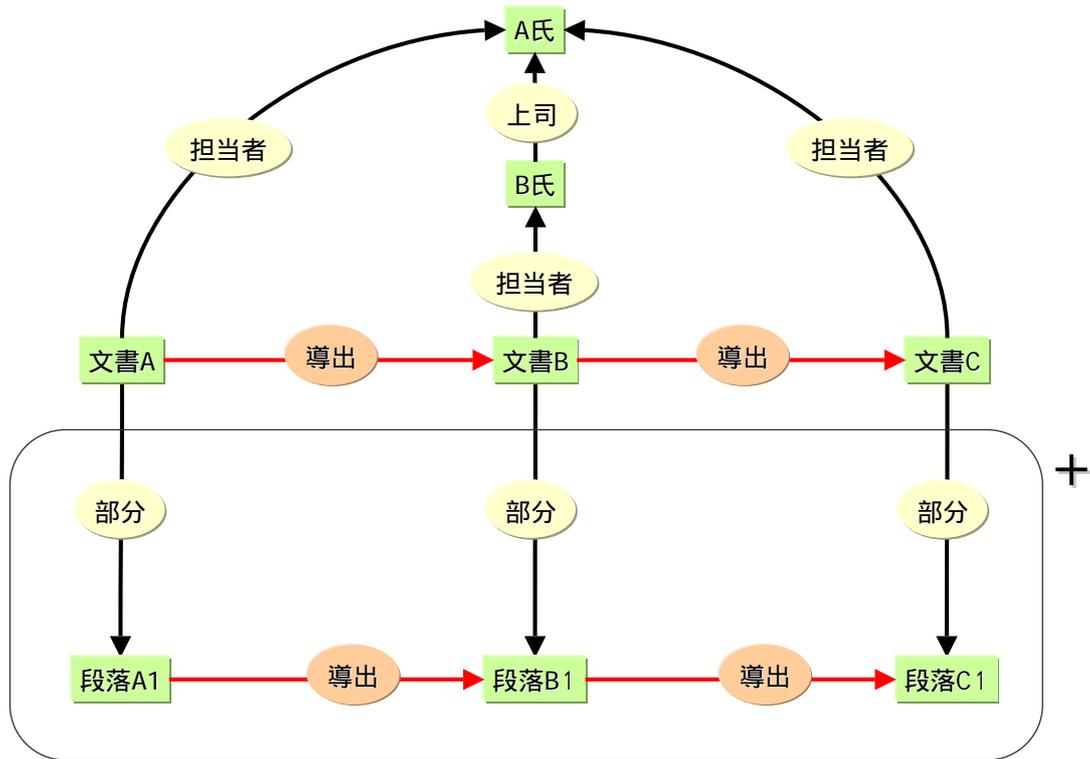
34

# セマンティックな業務システム

## ワークフローを表現するコンテンツ

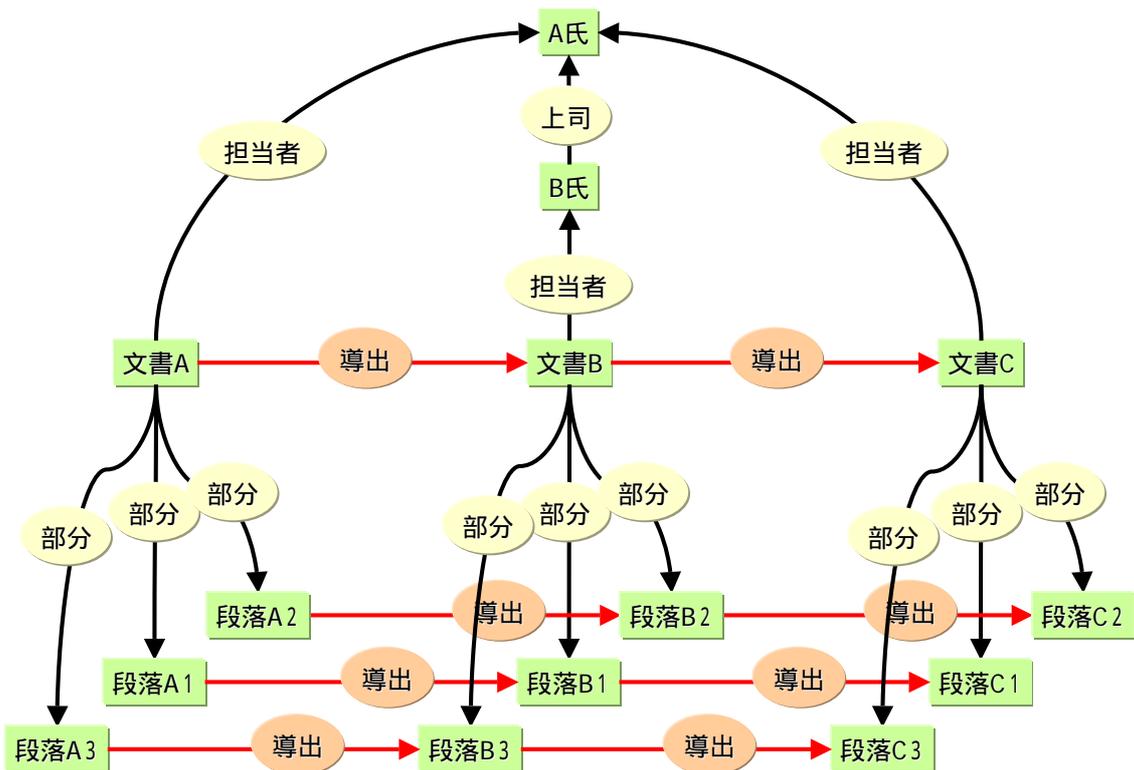


# コラボレーションの制約



37

# 制約が許容する構造



38

# 入力画面

## 山田 花子さんの基本チェック

煙草を吸う  はい  いいえ  
酒を飲む    
ストレスを感じる    
他にあれば入力して下さい

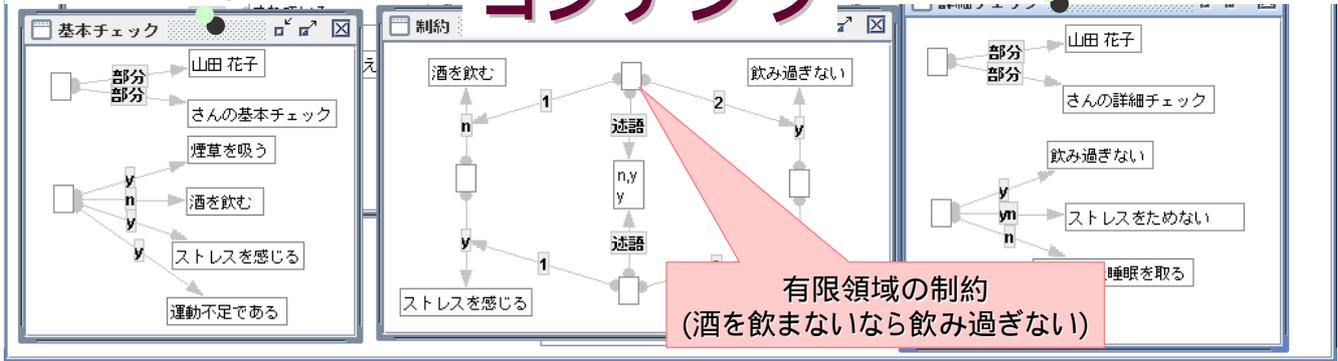
ウィンドウを閉じる 入力をキャンセル 過去の入力を修正

## 山田 花子さんの詳細チェック

飲み過ぎない  はい  いいえ  
ストレスをためてない    
7時間以上睡眠を取る    
他にあれば入力して下さい

ウィンドウを閉じる 入力をキャンセル 過去の入力を修正

# コンテンツ



# 現場の能力を生かす業務システム

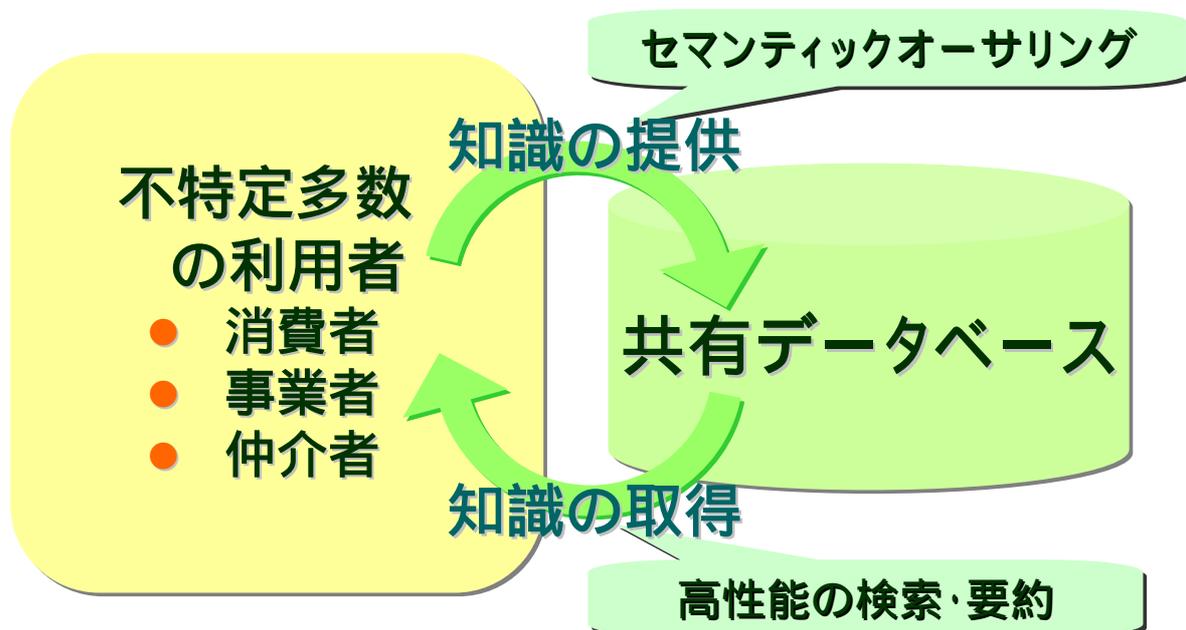
- 業務の意味に基づいてシステムを設計・運用
- 業務の設計が見える・触れる
- 漸進的・創発的な全体最適化
  - 利用者による改善の積み重ね
  - 業務と規程とシステムの擦り合わせ
- システム発注の明瞭化

# 知識循環型社会

## 知識循環

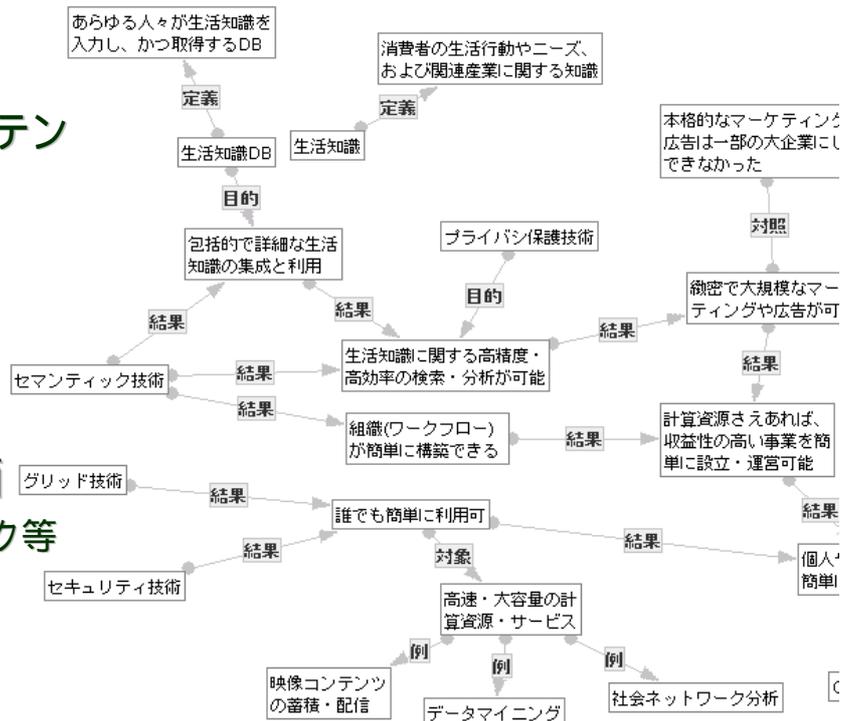
巨大なグループウェア

- データベースを作る人々 = 使う人々
- 知の社会的共有と拡大再生産



# 研究における知識循環

- 巨大なグラフ型コンテンツの共創
- 論文の廃絶
  - 論文でなく文を发表
- 知識循環の迅速化
  - 1週間?
- IF等より正確な評価
  - 人間関係ネットワーク等
- 理解しやすい提示
- 検索、翻訳、要約



# 循環型社会と全体最適化

- EA (Enterprise Architecture)
  - 情報システムを含む業務の統合的モデル化
    - cf. 状況オートマトン
  - 業務(だけ)の全体最適化
- 生産・消費・生活にわたる全体最適化
  - 職場・地域社会・家庭等にわたるコラボレーションの統合的モデル化と最適化
  - cf. 循環型経済: 地球環境を含めた最適化