

なんでもテーマ検索: Wikipediaの項目から情報探索リソースへの誘導

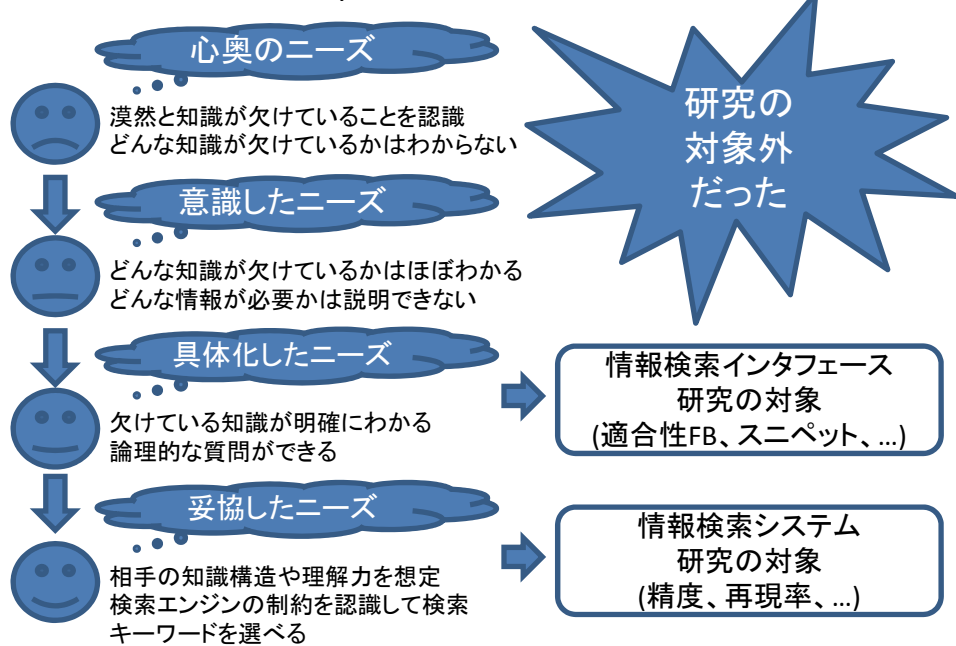
清田陽司(東京大)、田村悟之、坂井哲(東京電機大)

研究の目標

- 情報探索プロセスのうち、従来の情報検索システムに関する研究ではほとんど扱われてこなかった「情報ニーズの明確化」のプロセスの解明に取り組む
- 「情報ニーズの明確化」に適したオントロジー(分類体系)を構築する
- 「情報ニーズの明確化」のプロセスを評価するための方法論を確立する

情報探索のプロセス

情報ニーズの4段階 [Taylor, 1968]



情報探索の2つの世界

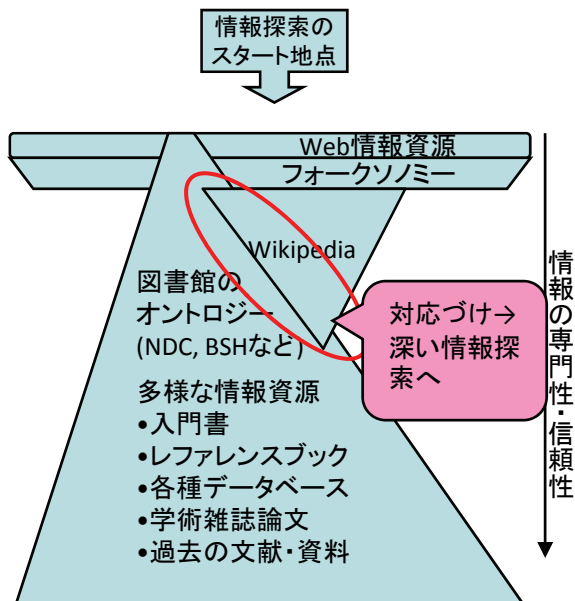
図書館

- ストック情報中心(過去の情報蓄積が充実)
- 膨大な情報の管理方法: 専門家によるトップダウンの秩序形成が中心
- 統一的なオントロジー
 - 書架分類法
 - 件名標目表

Web

- フロー情報中心(最新の情報が充実)
- 膨大な情報の管理方法: 一般参加者によるボトムアップの秩序形成が中心
- 多様なオントロジー
 - Webディレクトリ
 - フォークソミー

両者の世界をつなぐことができれば、あらゆるキーワードを切り口にさまざまな情報資源にアクセスすることができないか?



Wikipediaは両者の中間的な性格を有している

- Webの汎用的な情報資源としては最も組織化されている(Wikipediaカテゴリ)
- Web上の他の情報資源との親和性をもつ

まずはWikipediaと図書館オントロジーの対応付けに取り組む

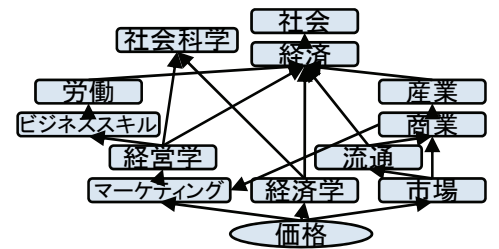
情報探索インフラとしてのオントロジー

書架分類法

- 1つの資料に1つのコード
- 国内では日本十進分類法(NDC)が広く普及
 - 000 総記
 - 100 哲学
 - 200 歴史
 - 300 社会科学
 - 400 自然科学
 - 500 技術・工学・工業
 - 600 産業
 - 700 芸術
 - 800 言語
 - 900 文学

Wikipedia

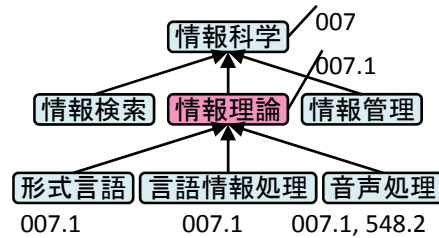
- 各記事に複数のカテゴリが付与されている(フォークソミーの特徴)
- カテゴリにもカテゴリが付与されている(ゆるやかな階層構造)



価格
価格(かかく)とは、有形・無形の各種の商品(サービスを含む)の取引に際して提示される金額を言う。基本的には需要と供給のバランスによって決定される。一般には、値段(ねだん)とも呼ばれる。
...
カテゴリ: [マーケティング][経済学][市場]

件名標目表

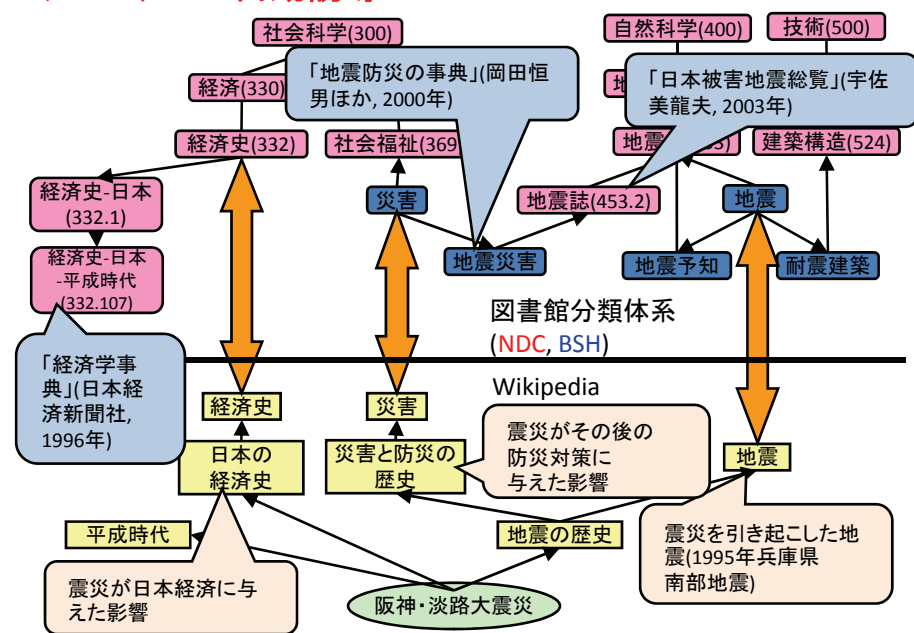
- 目録に付与する統制キーワード集合
- 国内では基本件名標目表(BSH)、国立国会図書館件名標目表(NDSLH)など



	件名標目表	Wikipedia
アプローチ	トップダウン	ボトムアップ
長所	深い組織化	新概念への対応
短所	新概念への遅れ	浅い組織化
上位概念	1つのみ	複数もてる
概念総数	11,184 (BSH4)	15,532
一致概念数		1,363

提案するアプローチ

Wikipediaカテゴリを媒体とした図書館分類体系(≒探索テーマ)への自動誘導



新たな情報探索インフラ実現の可能性

- 図書館の情報整理体系とWikipediaの統合利用
 - 情報探索の多様な切り口(テーマ)を示す
 - 情報探索に方向性を与える(信頼できる情報資源への誘導)
- 図書館情報資源とWeb情報資源の統合利用
 - 電子図書館の実現?

今後の研究テーマ

- 実運用システムによる評価
 - 図書館向け情報探索支援システム Littel Navigator のログ評価
- Web情報資源とWikipediaの対応づけ [田村2007]
- 類似したカテゴリ名の対応付け
 - 阪神・淡路大震災 ⇔ 地震災害
 - 日本の経済史 + 平成時代 ⇔ 332.107