

産業日本語研究会シンポジウム
2011年3月2日

知識インフラと日本語

国立国会図書館長
長尾 真

データから知識へ

- データ:加工されていない生の記録
取得における条件が明らかであることが大切
- 情報:データが何らかの文脈で解釈(理解)されたもの。それぞれの集団によって共通の意味を汲みとられる。

- 知識：情報を秩序化、体系化、抽象化し、他の知識との関係性を付けたもの。
- データや情報の解釈（理解）に必要なもの
- 知：知識を超えた、慣習や善・徳に支えられた判断をともなう何ものか（？）

図書館における知識の組織化

- 分類: 10進分類法、NDL分類法、……
- 主題分析: 資料の内容を分析し、定められた概念語を幾つか与えること
内容を表現する概念語はあらかじめ定めておく(シソーラスを構成)
- 典拠ファイル: 著者名、団体名、……
- シソーラス
- リサーチ・ナビ

情報を得たものが勝つ

- 情報を得ることの大切さ
- 単に情報が大切なのではなく、そこから何を導き出すかが重要である
- 今日の情報の量は爆発的であり、人手ではそれを詳しく分析することはほとんど不可能である

日本は情報にうとい

- 世界各国は情報を分析するチームを持って、情報収集に大きな努力をしている
- 意図的に情報を発信するなど、情報戦争もしている??
- ネットワークを介して相手の情報を破壊したり、盗んだりが行われている??

情報力

- 情報の収集力
- 情報の分析力
- 情報の価値化力
- 情報の発信力
- 情報の知識化・文化化力

情報分析(1)

- 世界各国で発信されている膨大な(草の根)情報の収集
- 自国語に翻訳し、整理・分類し、要約して提示
- 収集した情報の信頼性、信憑性の推定

情報分析(2)

- 分類(要約)した情報相互間の関係性を調べる
- 特に相反する(矛盾する)情報の存在を検出する
- ロングテイルにある情報で重要な情報の発見に努める

情報の時系列分析など

- あるテーマについての情報の時系列的变化を追跡して分析・提示する
- 新しい情報・事象を発見する
- 情報の地域的分布状況を調べる

知識インフラの構築

日本の研究の優れた点・弱い点

- 要素技術の研究は突出
- これらの成果は日本企業よりは外国ベンチャー企業で多く使われているようである
- 日本は要素技術を組み合わせ、システム化し、実用にもってゆくのが下手

イノベーションを興す要件

- 異分野の科学技術を組み合わせ、融合させ新しい科学技術分野を作る
- 異分野の科学技術を組み合わせ、システムとして統合し、新しい展開をはかる
- こういった努力は研究ではないといった偏見を持つことに大きな問題がある

経済・社会にインパクトを与え 価値を創出する研究開発

- 科学技術分野だけでなく、人文社会科学的視点が必要。学術全般がかかわる
- システム的視点に立ち、様々な環境を考えたグローバルな立場からの研究開発が大切
- これからは課題解決型の研究が大切であり、社会の人達が参加することが必要となるものが多い

知識インフラの必要性

- 知識の拡大再生産のためには、知識の創造と集積・流通・活用のサイクルの構築が必要
- 課題解決型の研究には様々な学問分野がかかわるシステムのアプローチが必要
- 課題を設定するためには、その課題についてこれまでどのような研究がなされて来たか、何が未解決か、イノベーションをおこせる可能性があるか、社会に対するインパクトはどうなりそうか等を調べねばならない

知の共有化

- 多くの分野がかかわるシステムの課題の場合、理工系の研究者だけでなく、政策立案者、人文社会系の研究者や市民もが調査してアセスメントができる環境を作る必要がある
- あらゆる学問の成果、企業社会、人間社会、自然社会等の知識・情報を収集整理し、自由に利用できるようにしなければならない

知識インフラの構造

- 研究情報基盤の整備が謳われてきたが、通信ネットワークが中心
- 必要なものは学術情報コンテンツ、知識コンテンツの組織的な整備
- 分野を超えた知識の関連付け
- 日本中に散在するコンテンツの所在を集中管理し、そこに検索をかければ関連する全ての必要なコンテンツが得られるようにする

- 知的基盤というよりは知識を取り出す知識基盤の構築
- 知識は関連するものが有機的に結合され、ネットワーク的に統合化されたもの(単に情報を集めたものではない)
- 日本中にある人文社会科学を含んだあらゆる学問・研究のコンテンツ、数値データ、研究データ、研究ツール、社会状況データ等が知識の形に組織化される必要
- 諸外国の同様なシステムとリンクがとれる必要

言語処理技術の重要性

言語情報処理(1)

- インターネット上の情報の言語横断的検索
- 検索結果の日本語への翻訳(多言語機械翻訳)
- 類似情報を一つの塊に集める(クラスタリング):
テキストの類似性検出
- クラスタされたテキストの要約

言語情報処理(2)

- テキスト中の特定の語句(テロ、・・・)の存在の検出
- テキスト中の記述部分と意図部分の分離、心理的・感情的部分の抽出
- テキスト間の類似性、関連性によるハイパーテキスト構造の形成

情報内容の信頼性チェック

- 情報内容の学問的知識との整合性チェック
- 相矛盾する情報が存在するかどうかのチェック
- テキスト表現が稚拙かどうかのチェック
- テキスト中の発信者の意見部分が断定的か推定的かなどのチェック
- テキストに発信者の名前が明記されているかどうかのチェック
- 多くの人が同一の意見であるかどうかのチェック

得られた情報から推定できること

- 既存の知識と得られた情報から推論して言えることを示す
- この推論された情報と得られた情報との間に矛盾するものがあるかどうかのチェック
- この情報はどの程度信頼できるか
この情報は正しいかどうか

を示すときには、その理由も示すことが大切である

情報分析のための基礎学問

- 情報分析研究は、テキスト等(テキスト、音声、画像、映像など)のアーカイビング、テキスト等の検索、重要語等の抽出、機械翻訳、クラスタリング、記号論理、推論、知識の蓄積と活用、文字等の認識などの多様な技術の基礎の上に行われる総合的な奥の深い興味のある学問分野である。