

誰もが安心・安全にWebコンテンツを利活用 できるための情報信憑性検証技術の実現

菱沼 宏之

総務省 総合通信基盤局 電気通信事業部 電気通信技術システム課

平成20年1月30日

情報の巨大集積化と利活用基盤技術開発連携群の活動

～情報爆発時代に果たす日本の役割と連携強化～

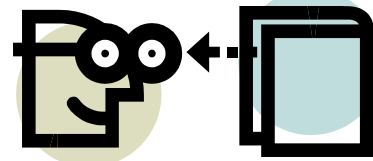
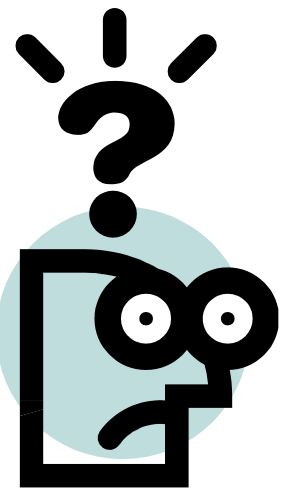
シンポジウム

なぜ情報分析が必要なのか

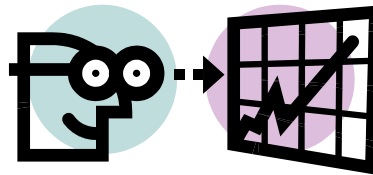
インターネットが一般的でなかった頃・・・



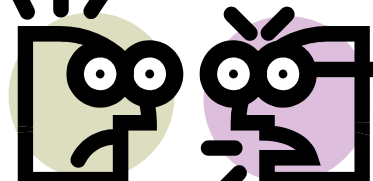
何かわからないことがあると・・・



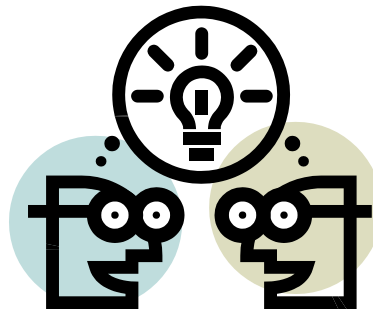
信頼できそうな本を調べ



異なる分野の情報を調べ

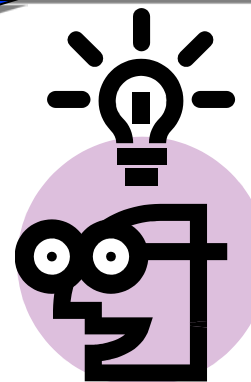


専門家の意見を聞き



他人の評判を聞き

発信元(源)の質に敏感となって
内容を総合的に判断した結果



調べた情報の質も理解した上
で、疑問に対する回答を自ら
導きだす。

現在は？

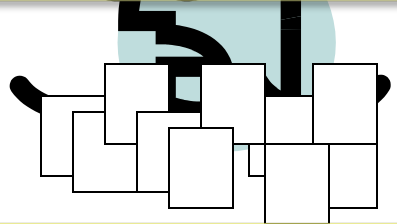
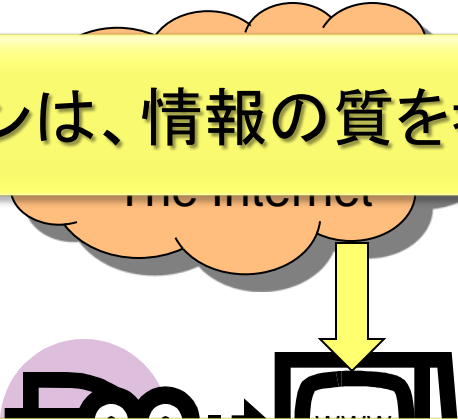


既存の検索エンジンは、情報の質を考慮しているのか？

大量の情報のため情報の質を判断することが出来ず



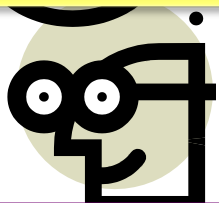
何かわからないことがあると...



既存の検索エンジンは、ユーザの意図を理解してコンテンツを検索しているのか？



思いついたキーワードを検索エンジンに入力



既存の検索エンジンは、キーワードにヒットするコンテンツの存在を明らかにするものにすぎなくなっている

したがって、

検索エンジンは、キーワードにヒットするコンテンツがあるかないかを調べるには有効である。

しかし、

★トップランクのコンテンツですら、いつも価値のある、また信頼できる情報を含んでいるとは限らない。

★ユーザの意図を理解してコンテンツを探しているとは限らない

既存の検索エンジンでインターネットから価値のある情報を取り出すためには、

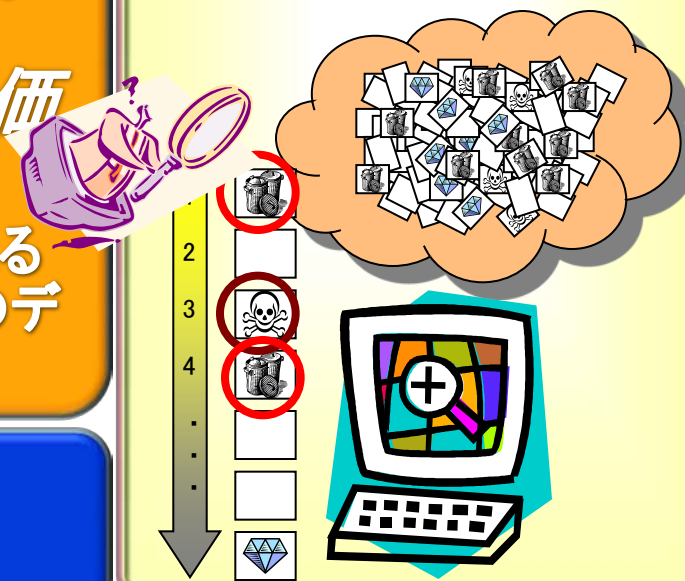
★ユーザは(検索意図を反映するために)何度も異なるキーワードで検索し、(質を判断するために)複数のデジタルコンテンツを閲覧して比較するしかない。

情報検索技術に加えて情報分析技術が必要となる。

既存の検索エンジン

現在の検索エンジンはメジャーなコンテンツを高く評価する。

サーチエンジンはキーワードで検索して、検索リストを出力する。(ユーザはトップランクのページに価値があると信じている。)

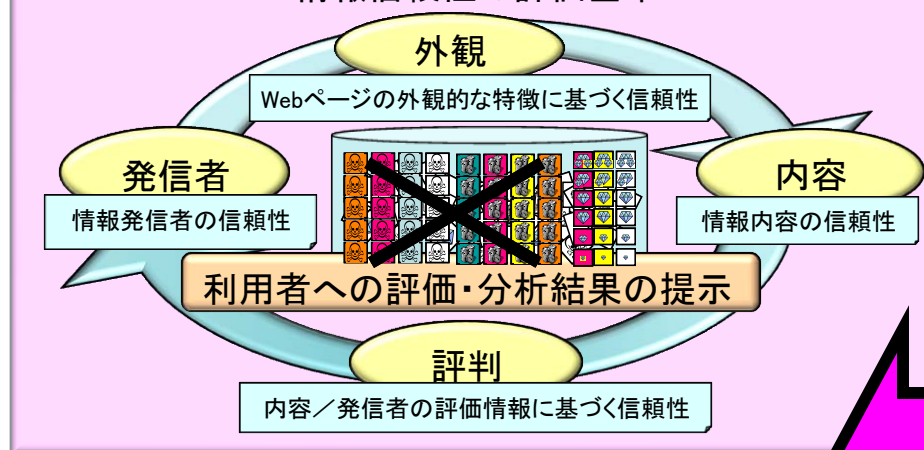


情報信憑性検証技術実現のための研究体制

情報通信研究機構(NICT)自ら実施する研究と、委託研究により、強力に推進。

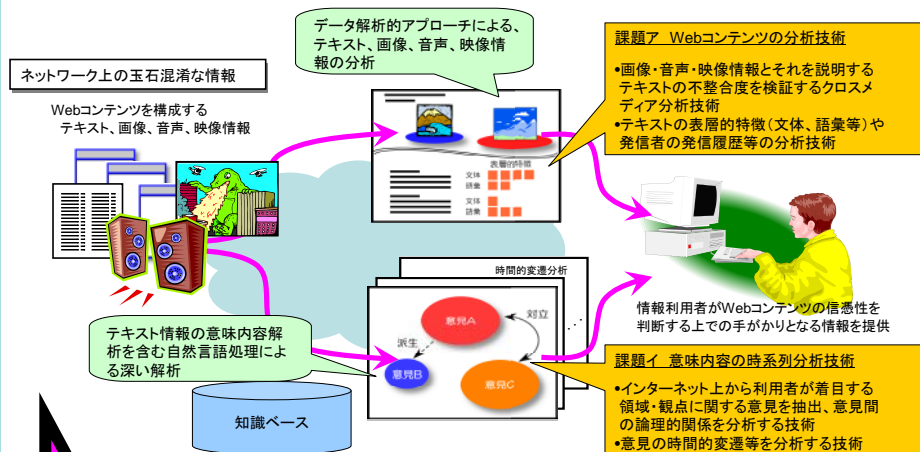
NICT研究

情報信頼性検証(情報の評価・分析・集約) 情報信頼性の評価基準



- 大量のWebコンテンツから自然言語処理技術に基づく4種類の情報分析手法で、ユーザが信頼性を判断する基準となる情報を提供

委託研究



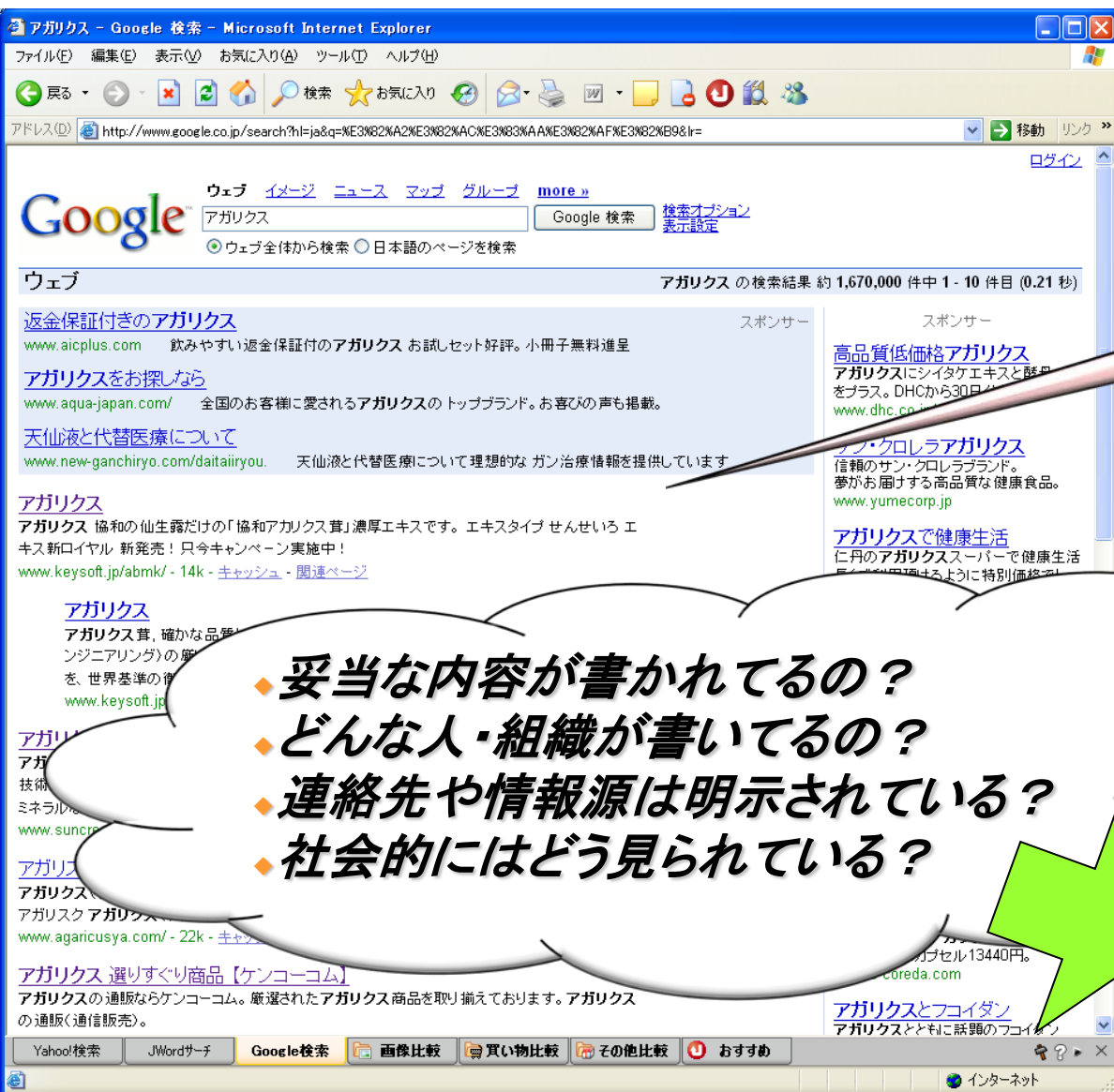
- データ解析的アプローチによるWebコンテンツを構成するテキスト、画像、音声、映像情報の分析技術
- 自然言語処理による深い解析による情報相互の論理的関係性、時間的変遷の分析技術

相互補完

誰もが安心・安全にWebコンテンツを利活用できるための情報信憑性検証技術を実現

信頼できる情報を見つけるには？

「アガリクス」の検索結果



健康によいという宣伝ばかり！
でも、本当に効くの？

信用するためには
いろいろな面から見たい！

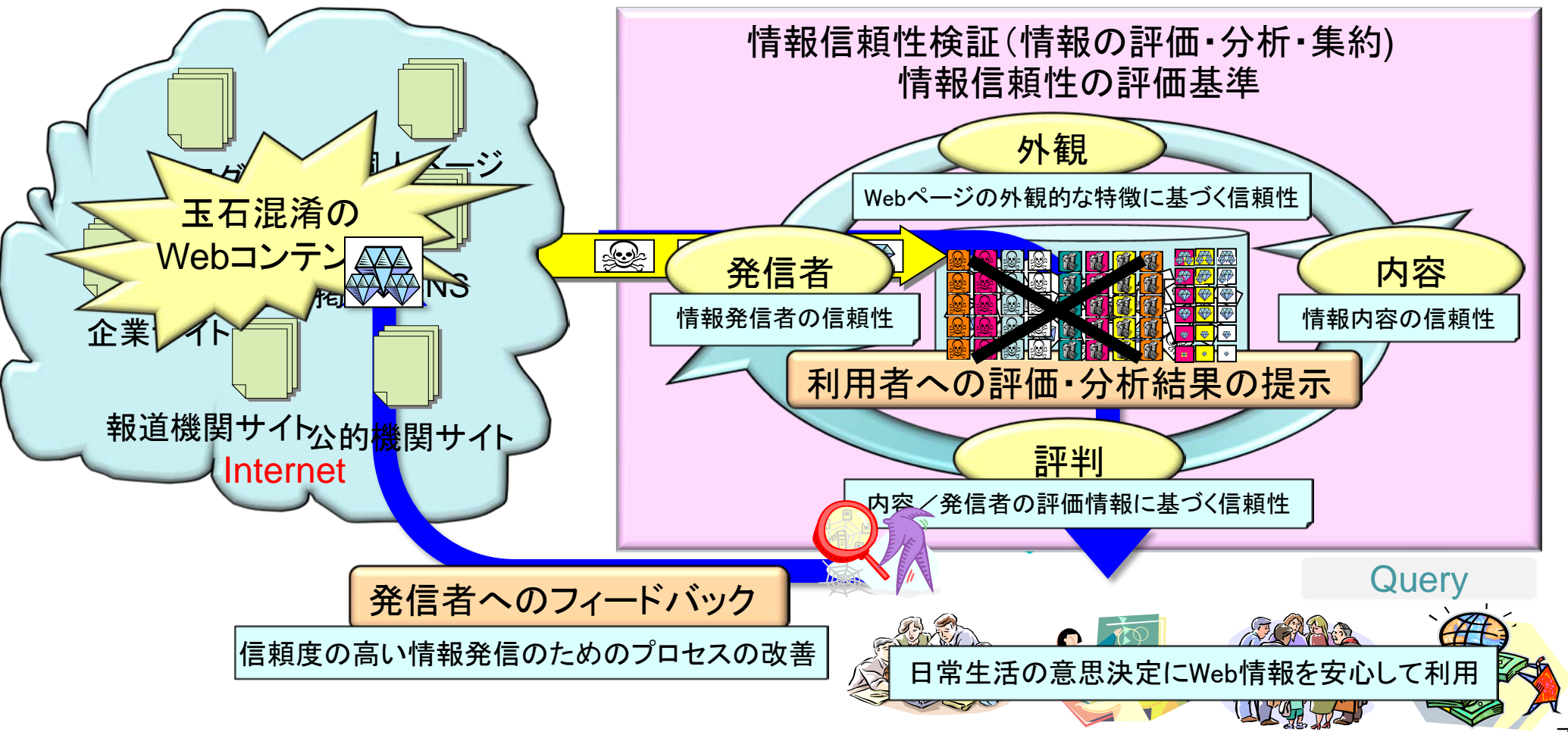
- ◆ 妥当な内容が書かれてるの？
- ◆ どんな人・組織が書いているの？
- ◆ 連絡先や情報源は明示されている？
- ◆ 社会的にはどう見られている？

既存の検索システム(グーグル等)
とは異なる
情報分析システムが必要！

自然言語処理技術に基づく4種類の情報分析手法で解析

- 情報通信研究機構(NICT)自ら研究開発するWISDOM*(情報信頼性検証システム)は、大量のWebコンテンツから自然言語処理技術に基づく4種類の情報分析手法で、ユーザが信頼性を判断する基準となる情報を提供する。

【*WISDOM:Web Information Sensibly and Discreetly Ordered and Marshaled】



WISDOMの画面例

情報発信者の分析

分析方法
選択
(ここでは発信
者分類を選択)

入力キーワード

情報の外観と評判の分析

Webページの要約

発信者の情報

外観の特徴

意見対象

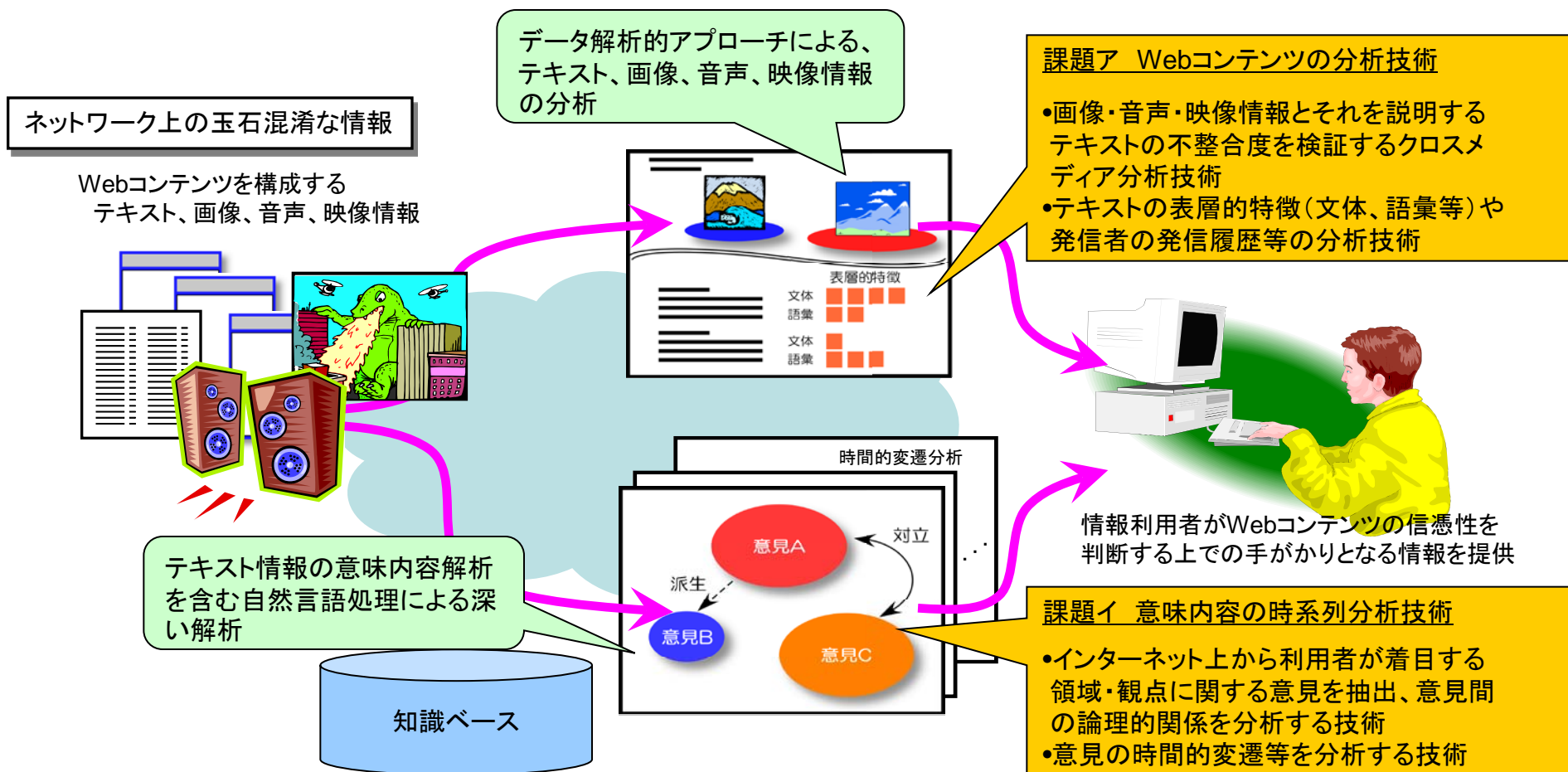
ページ中の意見と極性

The screenshot displays the WISDOM web application interface. At the top, there is a search bar with the keyword 'アガリクス' (Agaricus) entered. Below the search bar, there are navigation tabs for '発信者分類' (Sender Classification), 'ページ分類' (Page Classification), '概念分類' (Concept Classification), '意見' (Opinion), '意見根拠' (Opinion Basis), 'Q & A', and '絞込条件' (Filtering Conditions). The '発信者分類' tab is selected, showing a list of sender categories on the left, including '個人(36)', '有識者・専門家・著名人(11)', '一般人(4)', '匿名・ハンドルネーム(21)', '団体(62)', '富利団体(45)', '企業(45)', '業界団体(0)', '非富利団体(6)', '行政(3)', '公益法人等(2)', '大学(0)', '学会(0)', '任意団体(1)', '報道機関(11)', '新聞社(11)', '雑誌(0)', 'テレビ・ラジオ(0)', '個人の集合(1)', '参加者は実名(0)', '参加者は匿名(1)', and 'その他(0)'. The main content area shows search results for 'アガリクス', with the first result titled 'アガリクス、プロポリスの抗がん効果と安全性' (Agaricus, Propolis's Anti-cancer Effect and Safety) from 'キャッシュュ' (Cashu). The second result is titled '「アガリクス」発がん促進、キリン子会社4品を回収' (Agaricus Promotes Cancer, Kirin Subsidiary 4 Products Recalled) from 'キャッシュュ'. The third result is titled 'アガリクスで発がん促進' (Agaricus Promotes Cancer) from 'キャッシュュ'. The fourth result is titled '健康食品、2社は安全' (Health Food, 2 Companies Safe) from 'キャッシュュ'. On the right side, there is a sidebar with '情報発信者' (Information Sender) showing '中島久美子@医療情報部@株式会社読売新聞' (Kumiko Nakajima @ Medical Information Dept @ Yomiuri News Corp) and '情報外観' (Information Appearance) with buttons for '連絡先-実世界' (Contact - Real World), 'サイト内検索' (Site Search), 'Privacy Policy', '複数言語' (Multiple Languages), 'レイアウト' (Layout), 'サイト内ナビ' (Site Navigation), '更新-3日以内' (Update - Within 3 Days), and '広告-やや多い' (Ads - Slightly Many). Below the sidebar, there is a section for 'このページに含まれる意見対象・属性' (Opinion Targets and Attributes on this Page) with a bar chart showing '健康食品のマイナス情報(否定)' (Negative Information on Health Food (Denial)) and a legend for '否定' (Denial) and '肯定' (Affirmation). At the bottom, there is a section for 'ページ中の意見と極性' (Opinion and Polarity in the Page) with a bar chart showing '否定' (Denial) and '肯定' (Affirmation).

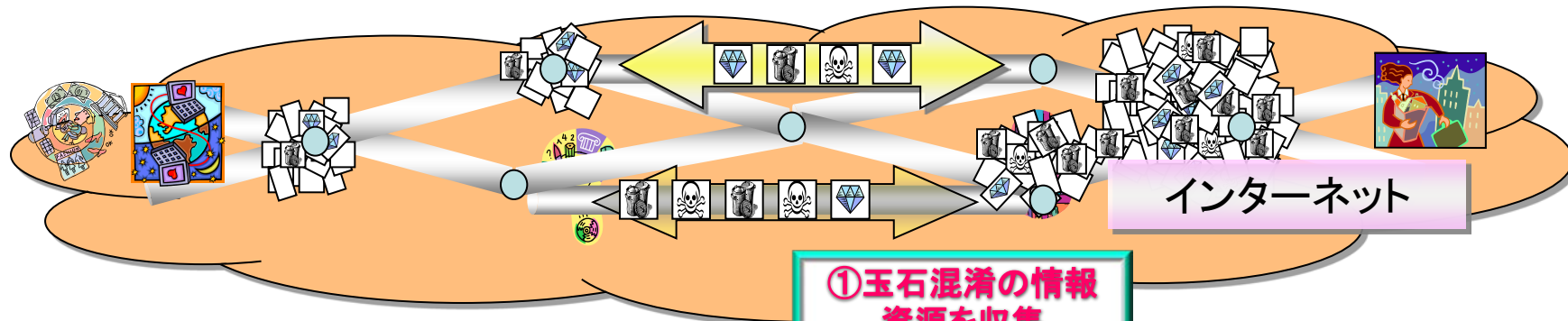
発信者に基づく
分類項目

選択した分類項目
のページリスト

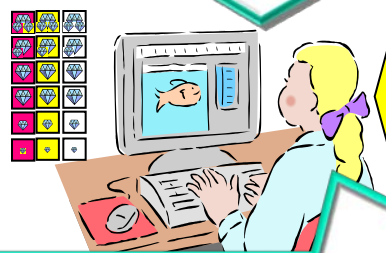
- データ解析的アプローチによるWebコンテンツを構成するテキスト、画像、音声、映像情報の分析技術
- 自然言語処理による深い解析による情報相互の論理的関係性、時間的変遷の分析技術



利用イメージ



【利用例】アガリクスって本当に健康に良いのかしら？WISDOMを使って検索エンジンの検索リストをみると、リスト上位は販売会社ばかりよね。公的機関の内容を見る限りでは健康によいのかどうか疑問のようね。アガリクスに関して多面的に検索をしてみよう。

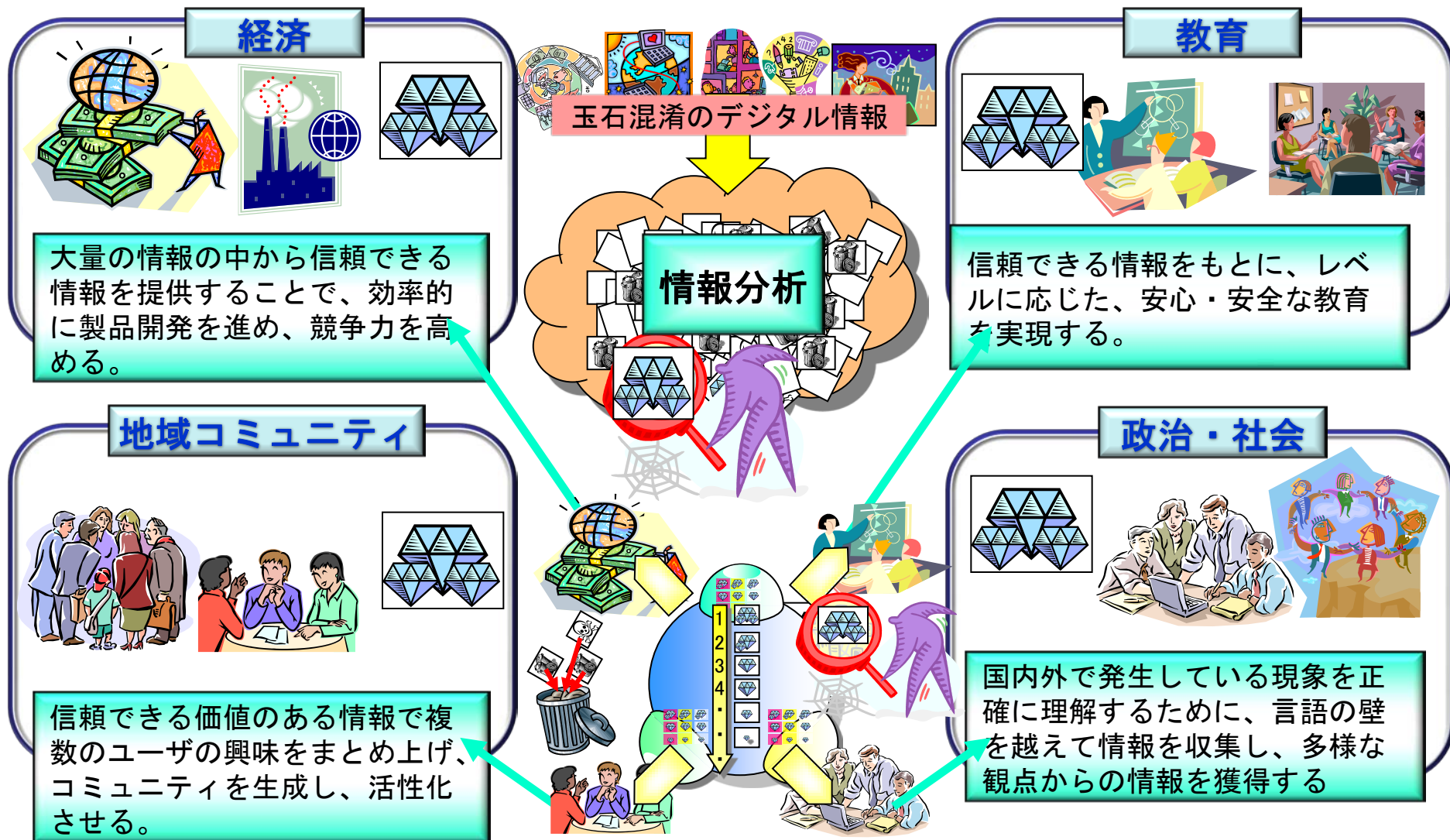


③分析結果を提示, ユーザにインターネット上の価値のある知識情報を発見させる。

②玉石混淆の情報資源を分析

アウトカム

信頼できる情報を正しく利活用できる情報通信基盤を構築し、ユビキタスネット社会への革新を加速させる。



- **現在のWeb**
 - インターネットによる情報の集合
- **次世代のWebとは？**
 - 信頼できる情報の利活用によるユビキタスネット上の「知の還流」を実現
- 以下の技術の研究を通して次世代Webへの変革を目指す。
 - 電気通信サービスにおける情報信憑性検証技術
 - 信頼できる情報を発見するための情報分析技術

ご静聴ありがとうございました

