

研究テーマ：情報コンテンツの信頼性評価プロジェクト(1/2)

(プロジェクト番号 JGN2-A18015)

研究機関： 情報通信研究機構知識創成コミュニケーション研究センター
知識処理グループ

研究の概要：

計算機・計算機ネットワークの進展に伴い、ウェブ、電子メール、企業内文書 など、言語によって表現される情報・知識の流通・集積が膨大となり、これが人々の生活の判断基準を与え、国家の政策、企業の経営方針までも左右しはじめている。このような文書情報の中から、信頼できる情報を取り出し、特定の課題についての情報を組織化して、背景的知識、事実、論点、意見分布などを 的確に抽出する技術は、今後の健全な社会のための基盤技術となる。本プロジェクトでは、このような情報信頼性評価に関する総合的な研究開発を行う。

研究の目的：

- ウェブなど膨大な文書情報の中から信頼できる情報を検出・組織化
- 背景的知識，事実，論点，意見分布などを的確に抽出する技術の研究開発

信頼できる情報を見つけるには？

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
いろんな面から見たい！

従来の検索とは異なる
新しい分析システムが必要！

- 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る？
- 社会的にはどう見られている？

- 妥当な内容が書かれてるの？ → 情報内容の分析
- どんな人・組織が書いているの？ → 情報発信者の分析
- 連絡先や情報源は明示されてる？ → 情報外観の分析
- 社会的にはどう見られている？ → 社会的評価の分析

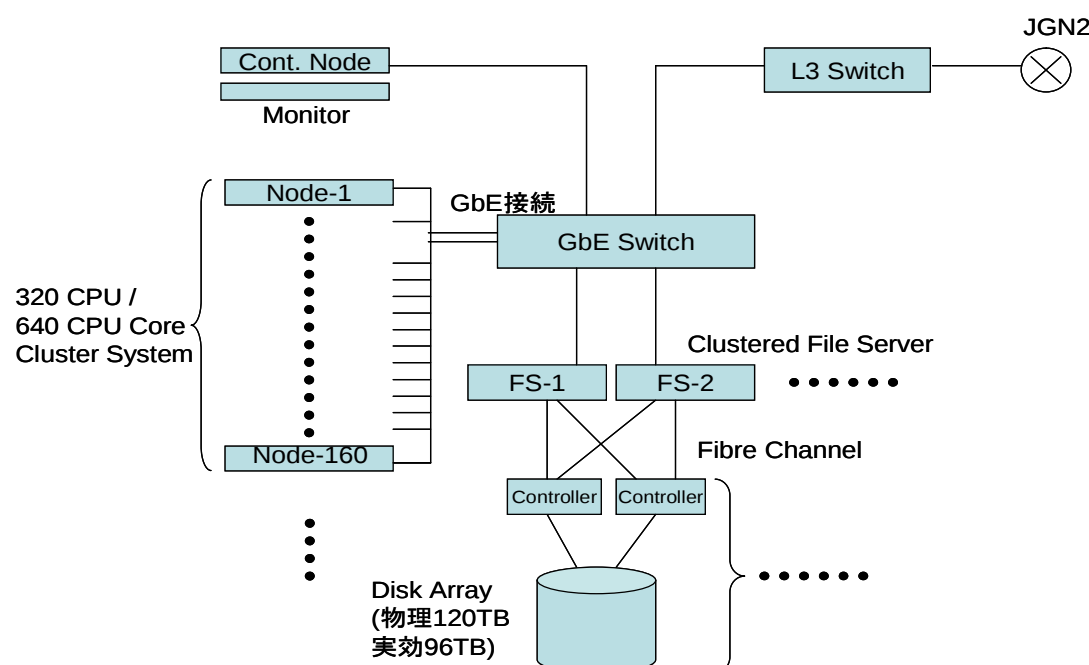
多角的な観点で情報を整理・組織化

さまざまな角度で情報を俯瞰できる

信頼できる情報の発見！

実験機器構成：

- クラスタシステム
(計算ノード200台、800 CPU Core)
- クラスタ型ファイルシステム
(サーバ追加によりスループット向上が可能)
- 大規模ストレージ装置
(物理容量120TB,実効容量96TB)
- JGN2と接続



情報コンテンツの信頼性評価プロジェクト(2/2)

(プロジェクト番号 JGN2-A18015)

研究機関：情報通信研究機構知識創成コミュニケーション研究センター
知識処理グループ

研究開発成果：

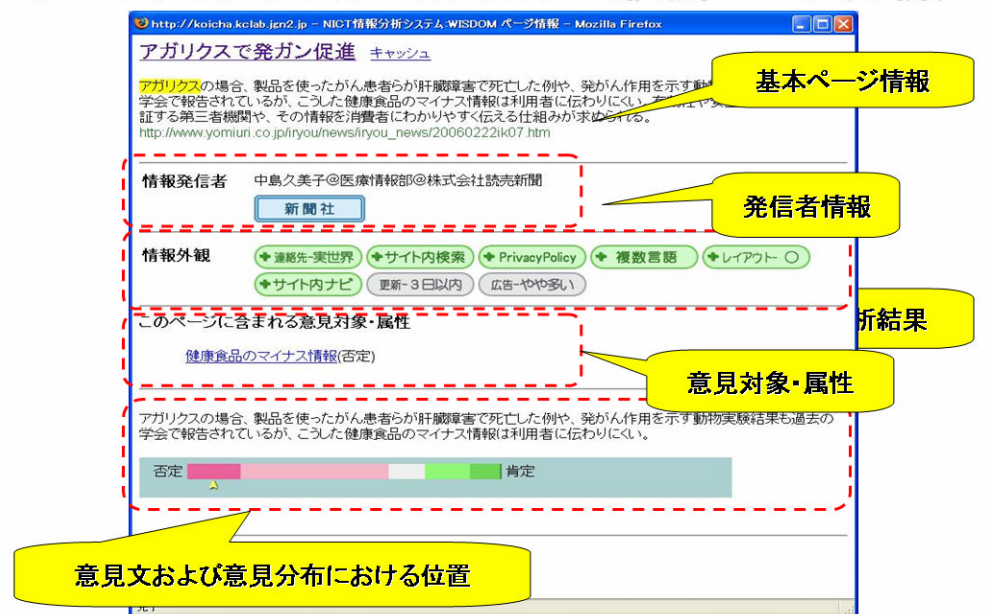
○情報分析システムWISDOMのプロトタイプ構築

○日本語Webコンテンツで現実的な規模に近い4億ページの収集を完了

○自然言語処理に基づく分析処理(現在までに、発信者属性(発信者と所属機関の関係記述等)、評価表現(暗黙的な評価も包含等)等の分析手法を開発。実験しつつ改良中

プロトタイプシステムWISDOM(ページ群分析)

プロトタイプシステムWISDOM(個別ページ分析)



プロジェクトのアピールポイント

本プロジェクトのポイントとして、インターネットから信頼できる情報をいかに見つけるかという課題に対し、ウェブ上のテキスト文書を対象として分析すること、また、分析にあたって、情報内容、情報発信者、情報外観、社会的評価といった4つの基準から行うこと、といった2点が挙げられる。本プロジェクトの成果である情報分析システムWISDOMを利用することで、利用者は、多面的な情報の評価軸をもつことになり、自分の興味のあるトピックについて、信頼できる情報をより確実に見極めることができるようになる。既存の検索エンジンではその検索結果は単一の評価軸で提示され、検索クエリによっては宣伝ページばかりが上位に表示され、われわれが本当に必要とする有用な情報が下位に埋もれることがしばしば起きる。本プロジェクトの成果はこのような情報の「質の壁」を打破する有効かつ強力なツールの一つとして一般の利用者に提供されるものであるといえる。

プロジェクトの自己評価

本プロジェクトは、当初の計画に基づき、研究基盤整備から具体的な最終ゴールの設定、要素技術の開発と改良と段階を追って着実に研究開発を実施してきており、プロジェクト管理は適切であると評価できる。