

## ドメインを考慮した論文の理解支援に関する研究

中村健二<sup>†</sup> 田中成典<sup>‡</sup> 大谷和史<sup>‡</sup> 栗田泰伸<sup>‡</sup>関西大学大学院総合情報学研究科<sup>†</sup> 関西大学総合情報学部<sup>‡</sup>

## 1. はじめに

近年、論文の電子化が進み、その入手が容易になった。入手した論文を理解するためには、論文の研究分野であるドメインについての知識を得る必要[1]があり、そのためには同様のドメインを扱った複数の論文を纏めたサーベイ論文が有用[2]である。サーベイ論文は、あるドメインにおける研究の流れを纏めた論文であり、ドメインについての包括的な知識を得るために有効である。しかし、あるドメインに属する重要な論文は相互に参照関係を持っているが、サーベイ論文がドメインに属する重要な論文を網羅的に纏めているとは限らない。そのため、サーベイ論文を閲覧するだけでは、ドメインにおける重要な論文を効率的に理解するための読解する順番を把握できるとは限らないという問題がある。この問題に対して、論文間の参照関係と関係属性から論文間の類似度を算出し、その類似度に基づき論文をドメイン毎にクラスタリングする研究[3]や、同じ論文を引用する共引用関係を考慮して類似論文を検索する研究[4]がある。しかし、これらの研究では、直接の引用関係を持つ論文のみを考慮しているため、ドメインを効率的に理解するための論文の読解順を把握することが困難である。そこで、本研究では、リンク構造解析の一手法である HITS (Hypertext Induced Topic Selection) アルゴリズム[5]を拡張し、サーベイ論文で紹介されている手法に関連する引用論文のドメインの流れとなる論文群を論文データベースから検出し、ドメインの流れを考慮した論文群を抽出する。そして、論文群を可視化することでドメインの理解に有用な論文群を提示する手法を提案する。

## 2. 研究の概要

本提案手法は、あらかじめ手作業で作成した

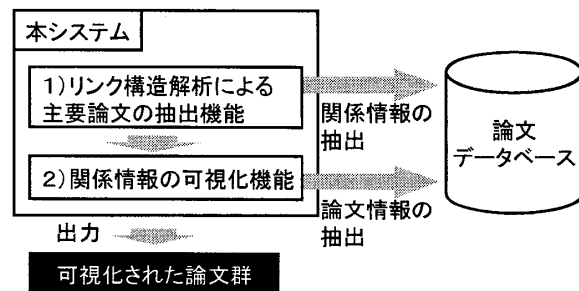


図1 システム概要

論文データベースを用意し、解析することでドメインの流れを可視化した図を出力する。本システム(図1)は、1) リンク構造解析による主要論文の抽出機能、2) 関係情報の可視化機能の2つの機能により構成される。また、論文データベースには、論文名、論文発表年、論文の引用関係を保存する。

## 2. 1 リンク構造解析による主要論文の抽出機能

本機能では、論文データベースに保存した論文に対して、発表年の差を考慮して拡張した HITS アルゴリズムを用いて主要論文を抽出する。解析結果から Authority 値が閾値以上である論文を主要論文として抽出する。本提案手法は、論文のドメインの流れを解析することに主眼を置くため、ドメインの流れの始まりとなる論文を取得する必要がある。ドメインの流れの始まりとなる論文は、常に新しい論文から参照されていることや、過去の論文が多いという特性があることに着目し、論文間の発表年の差を考慮して HITS アルゴリズムの拡張を行った。

## 2. 2 関係情報の可視化機能

本機能では、時系列順に並べた論文を可視化する。まず、リンク構造解析による主要論文の抽出機能で抽出した主要論文を引用している論文の中で、最も Authority 値が高い論文を取得する。次に、取得した論文を引用している論文の中で最も Authority 値が高い論文を取得する。以上の操作を繰り返すことでドメインの流れの上で、重要な論文を取得し、強調して可視化する。

Research for Understanding Support of a Paper Considering a Domain

<sup>†</sup> Kenji Nakamura,

Graduate School of Informatics, Kansai University, 2-1-1 Ryousenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

<sup>‡</sup> Shigenori Tanaka, Kazufumi Ohtani, Yasunobu Kurita  
Faculty of Informatics, Kansai University, 2-1-1 Ryousenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

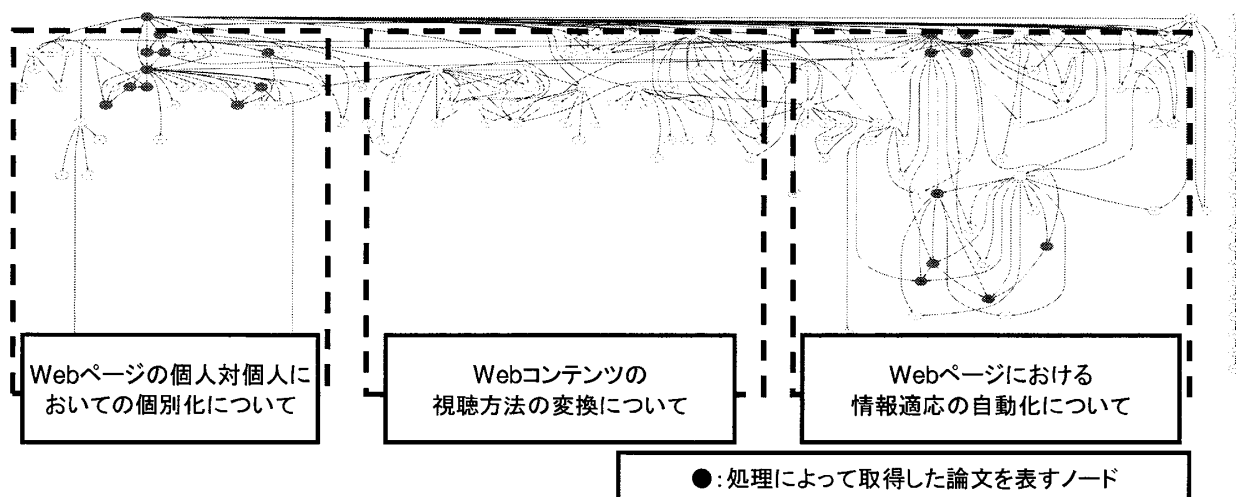


図2 システム出力

### 3. 提案手法の実証実験と考察

本システムの出力を図2に示す。提案手法の有用性を実証するため、本システムの出力を目視で確認した。

#### 3.1 実証実験

本実験で利用する実験データは、特定ドメインについて執筆されたサーベイ論文を基に収集した論文156件を使用した。使用した論文は、サーベイ論文の引用関係で結ばれている論文の中でも、Webからダウンロード可能なものを採用した。これらの論文情報をもとに手動で構築したデータベースを入力として、本提案手法にて得られた結果を可視化した図について考察を行った。

#### 3.2 結果と考察

本提案手法で作成した図から、ドメインの流れは、大きく3つに分類されることが分かった。また、その内の2つのドメインの流れに属する論文を強調して提示することができた。提示した2つの流れの内、Webページの1対1の個別化のドメインは、サーベイ論文が取り上げた手法における重要な論文についての流れを網羅的に提示できていることが分かった。またWebページの自動情報統合のドメインは、過去に発表された論文が重要であることから、過去からの論文の流れを正確に提示できていることが分かる。しかし、Webコンテンツの視聴方法の変換のドメインについては、流れを提示することができなかった。これは、引用数と被引用数が共に多い論文がドメイン内に存在しないことや、発表されてから時間が経過していない論文が多く含まれていたためであると考えられる。

#### 4. おわりに

本研究では、論文の理解に有用なドメインの

流れを考慮した論文の可視化を行った。サーベイ論文から引用関係で結ばれている論文を論文間の発表年の差を考慮して拡張したHITSアルゴリズムを用いて解析することで、ドメインの流れを考慮した可視化が可能となった。しかし、一部に同じ手法が使われているだけで本来は異なるドメインに所属している論文が同様のドメインに所属していると判断されることがあった。また、ドメインを対象とした理解支援を行っているが、手法や研究背景を考慮していないため、引用関係の解析だけでは取得することができない有用な論文を提示できていない。今後は、抄録や本文の内容を考慮し、手法や研究背景を考慮した論文理解支援を行う予定である。

#### 参考文献

- [1] 内田誠：ネットワーク構造から見た学術研究トピックの分析と学際研究の可能性発見，情報処理学会誌，情報処理学会，Vol.49，No.3，pp.72-73，2008.3.
- [2] 難波英嗣，奥村学：多言語論文データベースを用いたサーベイ論文検出，言語理解とコミュニケーション研究会研究報告，電子情報通信学会，Vol.102，No.199，pp.35-41，2002.7.
- [3] 難波英嗣，神門典子，奥村学：論文間の参照情報を考慮した関連論文の組織化，情報処理学会論文誌，情報処理学会，Vol.42，No.11，pp.2640-2649，2001.11.
- [4] White, D. and McCain, W. : Visualizing a Discipline: An Author Co-citation Analysis of Information Science 1972-1995, Journal of the ASIS, ASIS, Vol.49, No.4, pp.327-355, 1998.4.
- [5] Kleinberg, J. : Authoritative Sources in A Hyperlinked Environment, Journal of the ACM, ACM, Vol.46, No.5, pp.604-632, 1999.9.