

博士（人間科学）学位論文 概要書

**Web Information Search and Sharing:
A Human-Centric Integrated Approach**

人間中心の統合的アプローチによる
ウェブ情報検索と共有

2009年1月

早稲田大学大学院 人間科学研究科

シュティフ ロマン

Shtykh, Roman

研究指導教員： 金 群 教授

概要

現代社会において日々大量の情報が生産され、インターネットやイントラネットで共有されている。それらの情報の取り扱いについては、われわれは様々な問題に直面しており、とくにどうやって効率よく必要な情報を発見し再利用するか困難である。そのため、情報検索サービスの役割が非常に大きいと考えられている。しかし、われわれが経験している情報氾濫状況では、そのようなサービスは「検索問い合わせと保有ドキュメントのセット」しか考慮しておらず、迅速かつ正確に個人の検索ニーズに応えることができない。その結果、ウェブ情報検索の質を向上させるには、情報システムは各ユーザの情報ニーズを把握することが必要となる。しかし、ユーザのニーズを知り、的確に応用することは極めて難しく、多くの場合は不可能である。その一方、ユーザの複数の情報行動コンテキストを知ることにより、ユーザはある特定のコンテキストにおいてどういう情報を求めているか推定することができると考えられている。

われわれは、情報氾濫から生じる各種の問題を克服するために必要不可欠なヒューマン・ファクター、及び効果的なパーソナライゼーションを達成するためのコンテキストの重要性を認識し、ユーザ・システム間のインタラクションからコンテキストを的確に捉え、ユーザ間のコラボレーションのメリットを活かして、より良い情報検索・共有を達成するための有効なソリューションを探究する。

本研究の基本理念は人間中心であり、すなわち、ユーザを情報検索の中心に据えおくことである。われわれはユーザ・システム間のインタラクションから得られた適合性フィードバックより、ユーザにかかわるコンテキストを汲み取り、ユーザの興味を表すコンセプトとして組織化する。これらのコンセプトはユーザプロファイルの基礎要素としてユーザの検索エクスペリエンスをパーソナライズするために用いられる。また、コンテキストはユーザの興味や活動内容とともに変わるため、コンセプトの組織化の有効性を保つには、このようなダイナミズムを考慮しなければならない。さらに、われわれが提案するアプローチにおけるユーザ中心の特性を実現するためには、協調的な情報検索行動中で生じる「ソーシャル」な要素を活用する。本論文では、情報を探そうとしているユーザのニーズを「理解」し、よりよく満たすように試みる協調型情報検索パーソナライズ化の統合アプローチについて論じる。われわれが提案するアプローチは、いわゆる BESS (BEtter Search and Sharing) というウェブ情報検索フレームワークにより実現される。BESS フレームワークは上に述べたわれわれの基本理念と考えに基づいて構築されている。

本論文は、7つの章からなる。各章で述べられている内容を以下に概説する。

第1章は、序論である。情報氾濫の問題と課題を示し、情報獲得や処理への影響について論じる。そして、問題の主観的な特性を明確にし、その解決には社会的な側面を含むヒューマン・ファクターを十分に考慮する必要があると考える。

第2章は、情報検索や探索における人間中心のソリューション、特にパーソナライズ化、及びその特徴や問題について考える。そして、パーソナライズ技術に必要なユーザモデル化（ユーザプロファイル構築）の手法について述べる。

第3章は、ウェブ情報検索フレームワーク BESS について紹介する。本フレームワークは、ユーザから得た適合性フィードバックを利用してユーザの興味を推定し、それらの興味を反映す

るプロフィールを作成し、最終的に情報検索や共有に適用することによりユーザ中心の特性を実現する。本章では、提案アプローチの概念的基礎を論じ、フレームワークのモデルとアーキテクチャを説明する。

本研究では、ユーザコンテキストを推定するには、先ずユーザの興味を表すコンセプトを作成し、多層ユーザプロフィールに格納する。ユーザプロフィールは本フレームワークのコアであり、ユーザの興味変化とともに更新される。それによって、何時でもユーザの興味を正確に捉えられると考えられる。プロフィール生成と更新は、短期的興味の変わりやすさや長期的興味の安定性を反映する *recency*, *frequency* と *persistence* といった3つの基準要素を採用した「ユーザ興味変化駆動プロフィール生成」メカニズムによって行われる。第4章では、それらの説明を行う。

第5章では、ユーザプロフィールに用いられるコンセプトを生成する方法として、われわれは同一適合性フィードバックの特徴に基づいた *High Similarity Sequence Data-Driven* クラスタリング (H2S2D) を提案している。H2S2D は、データに関する予備知識を前提条件としないオンライン漸増手法であり、他の同類手法に比べて極めて迅速かつ実現しやすい。精度等を評価するため、さまざまな異なるデータセットを用いて実験を行い、その評価結果を示し、提案手法の特徴について論じる。

コラボレーションは、われわれが提案するユーザ中心の統合アプローチにとって必要不可欠な要素である。BESS では、ユーザの協調により、個人要素を含めた個人経験による付加価値が与えられた主観的 *subjective document index* が作成され、検索対象となる。第6章では、ユーザプロフィールの情報からユーザの経験値をはかるために *contribution activeness* と *contribution popularity* を提案し、ドキュメントのアセスメントの方法を示している。さらに、各ユーザプロフィール間の類似を用いた *focused search* を説明し、実例を示す。

第7章は、本論文の結論である。本研究の主な成果をまとめ、今後の研究課題を示す。

本研究では、ウェブ情報検索と共有のための人間中心の統合的アプローチを提案されている。われわれの提案アプローチは、ウェブ情報検索に、ユーザの個人的なコンテキストや、ユーザによる協力的な貢献と協同評価により実現される「ソーシャル」的なファクターなど、ユーザ中心の理念にとって大事な要素を取り入れられていることが特徴である。本研究の主な貢献は以下にまとめる。

- 1) より良い情報検索と共有のフレームワークを提案している。
- 2) ユーザのコンテキスト情報を的確に組織化するためのダイナミック興味変化駆動モデルを考案した。
- 3) コンセプト構築のための *High Similarity Sequence Data-Driven* クラスタリング・アルゴリズムを開発している。
- 4) われわれの提案するモデルとアルゴリズムを評価するための指標を新たに定義に、導入している。

提案アプローチは、概念的なレベルにおいて従来のパーソナライゼーション・ソリューションとは次の点で異なる。

- 1) 従来の検索サービスにおける個人的な要素が排除されたいわゆる「客観的な」データから、ユーザの過去の経験による付加価値のある「主観的な」データを分離され、そ

れによって、主観的な情報の蓄積と検索プロセス中での共有のための「ソーシャル」作業空間が生成される。

- 2) 以下に示すパーソナライゼーション・サイクルにおける情報共有と検索の統合が繰り返されている。

検索行動を通じた貢献（適合性フィードバック） → それを使ったユーザ
プロフィール作成 → ユーザプロフィールに基づいたフィードバックのアセ
スメント → 個人とコミュニティのコンテキストに基づいた検索結果ラン
キング → ……

パーソナライゼーション・サイクルの有効性と期待されるパーソナライズの質は、適合性フィードバックの類似に基づいた H2S2D オンライン漸増クラスタリング・アルゴリズムや「ユーザ興味変化駆動プロフィール生成」メカニズムなどにより確保されている。H2S2D は、同一適合性フィードバックの特徴を活かし、比較的単純である一方、適切なクラスタリング結果や速いクラスタリング速度が得られ、常に更新されたコンセプトの状態を維持することが可能となる。それに加えて、「ユーザ興味変化駆動プロフィール生成」メカニズムは、ユーザの興味変化に連動した適切かつ正確なコンセプトの組織化が保証され、ユーザにかかわるコンテキストの変化について常に正確な情報を提供することが可能である。さらに、評価実験ではわれわれが提案するウェブ情報検索と共有のための人間中心の統合アプローチの有効性と有用性を示されている。また、BESS はわれわれの協調型検索と共有の基本考えを実現する総合フレームワークであり、それを構成する基本手法や基盤技術が他の関連研究分野にも適用できる。

本研究の成果はウェブ情報検索分野全般、特にコンテキスト指向パーソナライゼーションの研究発展を促進し、大いに貢献することが期待される。