

2007 年度未踏ソフトウェア創造事業

# Search People on the Web with Privacy Preserving

## 成果報告書

開発者代表： 楊 征路

担当 PM： David J. Farber

プロジェクト管理組織： GLOCOM

2007年12月10日

# 目次

1. 要約.....	1
2. 背景及び目的.....	2
3. プロジェクト概要.....	2
4. 開発成果の特徴.....	3
4.1. 自由自在にウェブ上で人を検索.....	4
4.2. 個人情報の編集に関して柔軟になる.....	4
4.3. 個人のプライバシーを完全に守る.....	4
4.4. 個人向け情報システムに対する情報源となる.....	4
5. 開発者名.....	5

## 1 要約

情報を検索・ブラウズするとき、WWW はユーザにとって簡単だが効果的な手段である。近年、ウェブ上で有用な情報を探すための検索エンジンが普及されている。しかし残念ながら、現用の検索エンジンでは、人という検索される可能性があり、もっとも重要なオブジェクトの検索および要約ができない。なぜなら、人の名前を常に通常の文字列とみなされるため、名前を表すキーワードと人を結びつける自然な手法が欠けているからである。さらに、これらの検索エンジンでは、要約ではなくランキング順のみでウェブページを提供されているので、ユーザが手動で個人情報を抽出するのに多くの時間を費やすかもしれない。

このプロジェクトでは、私たちは融合システムを開発しました。(それは、リアルタイムで、ウェブで個人情報を検索することができました)。プロジェクトで実装されたいくつかのサブ機能があって、すなわち、Web クローラと、ウェブページの分類と、パーソナルデータの抽出と、パーソナル情報の要約と、プライバシーの保護とグラフィカルユーザインタフェイスです。

本プロジェクトでは、自由自在に他の人を検索し、知ることができるようになるだけでなく、最も重要なのは、個人情報を生かし、ユーザ・コンピュータ対話、携帯電話サービス、ソーシャルネットワーク、オンラインデート、求人、オンラインショップ、教育・トレーニングなど、個人指向システムの有効性および効率を大幅に向上できるかもしれない。

## 2 背景及び目的

人々が他の人に興味を持っているので、ウェブで他の人に関して知るのは友人を通して他の人に関して知るのと同じくらい当然です。有効な巨大な情報量がオンラインであることで、WWWは、個人情報に要約するための肥よくで理想的な領域です。残念ながら、個人的な情報を抽出と要約する問題は、これらの検索エンジンがいつも普通なストリングとして人の名前を扱うので伝統的なキーワードに基づいている検索エンジン(例えば、Google)で有効に取り組むことができないで、その結果、名前を表すキーワードと人間の間で自然な関連関係を欠きます。そのうえ、伝統的なキーワードに基づいている検索エンジンが概要の代わりに別々のウェブページを返すことができるだけである検索した。

「人々を探す」というこのプロジェクトは、伝統的なページレベル(例えば、Google)からのパラダイム・シフトがレベルが探すオブジェクト(すなわち、オブジェクトとしての人々)として探す現れについて上で説明された私の発見だけに基づいていますが、精度、有効性、および効率問題における多くの既存のシステムの性能を向上させるために大いにユーザの認識された関心と目標を当てにする多量のアプリケーションについても基づいています。

## 3 プロジェクト概要

本ソフトウェアは以下の様々な機能をもつソフトウェアの構築を目標として開発された。

- PHPにより実装されたGUI
- 効率的なウェブクローラ
- 効率的なウェブページ分類
- 効率的な個人的なデータ抽出
- 効率的な個人情報要約
- 個人情報を編集することができる
- 個人的なプライバシーは保護される
- 個人情報は、効率的に保存されて、管理される

システムのフレームワークは図1で示されます。

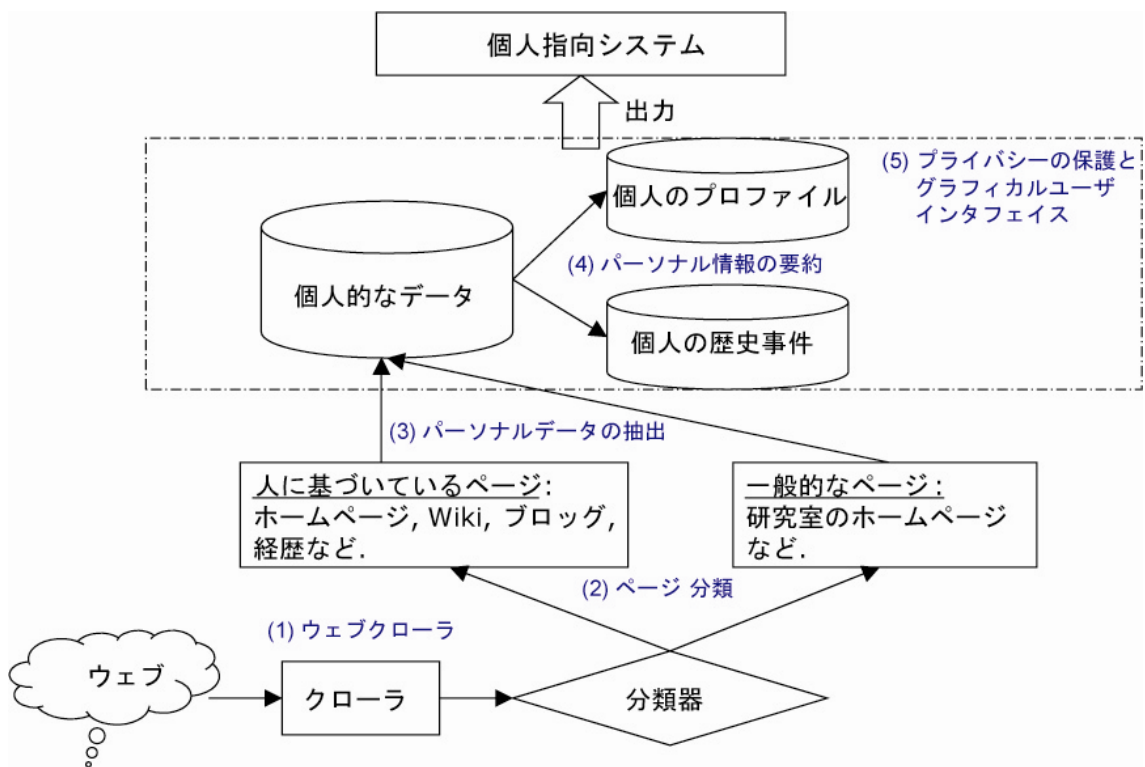


図 1. システムのフレームワーク

## 4 開発成果の特徴

### 4.1. 自由自在にウェブ上で人を検索

ウェブからの人々のためにそれを捜すのに開発されたシステムを使用することができます。人の名前を与えて、ユーザは妥当な時間中にその人の基準プロフィールと歴史イベントを見ることができます。言い換えれば、このシステムは見られて、これらのサーチエンジンに対する補完性があるかもしれません。有効で効率的な戦略は、個人情報を発見するためにシステムで工夫されます。

### 4.2. 個人情報の編集に関して柔軟になる

Wiki から評判システムのメカニズムがシステムで紹介されるのを学ぶのがそれがそれぞれのユーザ評判を与えるその人の過去の振舞いを基礎づけました。ユーザ(エディタ)は情報の正当性に投票します、そして、彼らの票は価

値の総スコアを数えるために蓄積されます。勝者は多くの票がある値であるべきです。

#### 4.3. 個人のプライバシーを完全に守る

その人がいくつかの賞を手に入れることができないなら、どんな人々もその人の個人情報をすべての他のものに発表したがっていません。このシステムでは、ユーザは、異なったデータを異種の他のグループ(サービス)に発表することによって、それら自身の個人情報の公開のレベルを制御することができます。

#### 4.4. 個人向け情報システムに対する情報源となる

このプロジェクトの出力は（個人情報）、ユーザ・コンピュータ対話、携帯電話サービス、ソーシャルネットワーク、オンラインデート、求人、オンラインショップ、教育・トレーニングなどに対して、個人指向システムの有効性および効率を大幅に向上できるかもしれない。

## 5 開発者名

楊征路（東京大学）