

## 2009 年度下期未踏 IT 人材発掘・育成事業 採択案件評価書

### 1. 担当PM

勝屋 久 PM(Venture BEAT Project 主宰)

### 2. 採択者氏名

チーフクリエイター:岡 瑞起(東京大学 知の構造化センター・特任研究員)  
コクリエイター:なし

### 3. プロジェクト管理組織

株式会社京王 IT ソリューションズ

### 4. 委託金支払額

2,800,000 円

### 5. テーマ名

ウェブデータからの行為抽出エンジンの開発

### 6. 関連Webサイト

<http://pingpong.ne.jp>

### 7. テーマ概要

Facebook、Twitterといったソーシャルネットワークサービスや、ブログやマイクロブログといったコミュニケーション手段の普及により、誰でも気軽に情報を公開、共有できるようになっている近年、さまざまなユーザが、製品やイベントに対しての嗜好

や意見、また、日常の行動などについて発信している。その結果、年々企業側が提供する製品に関する情報よりも、これらのコミュニケーション手段を通じてユーザが発信するコンテンツ(User Generated Contents、以下、UGCs)が、消費者の購買行動に影響を与えるようになってきていると言われている。このような背景の中、UGCsから自動的に注目の話題を抽出したり、製品やブランドに関する評判を分析するサイトやシステムが開発、公開されている。このようなさまざまなコミュニケーション手段を通じて発信されるUGCsから抽出可能な情報は、製品や出来事に対するユーザの意見、評判といったものに限らない。Twitterのサイトにおいて「いまなにしてる？」という問いかけに代表されるように、UGCsには多くの行為を含むデータが蓄積されている。

本プロジェクトは、ウェブから人々は何をしているのかを知ることを通じて人間の英知を集め、有効的に活用するシステムの開発を行う。具体的には、ウェブ上で、さまざまな人がブログやマイクロブログといった新しいコミュニケーション手段を通じて発信しているデータを収集し、これまでに開発・研究されてきているウェブマイニング技術と自然言語処理技術を用いて、誰が、いつ、どこで、何に対して、何をしたか、を集めた行為データベースを構築する。さらに、このデータベースを活用した行動パターンの可視化や、行動予測を行うウェブ上のサービスとして提供する。提供するサービスの情報を用いることにより、新しい情報環境のデザインなど、さまざまな分野のサービスをデザインするツールとして大きく貢献できると考える。

## 8. 採択理由

Twitter、Blog、mixi、Facebook などから情報抽出し、Webマイニング技術と自然言語処理技術を用いて人の行為を抽出するエンジンを開発することがテーマである。今回の開発をとおして、自然言語処理の要素技術で難しいとされている含意・言い換えの問題と事実性判定の問題を解き、大量の情報を解析するソフトウェア開発という高い目標にチャレンジをする岡さんの情熱と志を感じた。webマイニングの研究経験とGoogleやMicrosoft Researchなど海外でのインターンシップの経験もあり、グローバル視野をもつ人材としても今後期待できる。この期間中はTwitterにフォーカスにフォーカスして開発して、実用化を前提に取り組んでほしい。

## 9. 開発目標

「行為抽出エンジン:動く地図」システムを作成するにあたって、主に以下 4 つの機能の実装を行う。

- 1) Twitter からの行為情報の抽出
- 2) 建物内の行為情報を位置情報と共に取得するアプリケーションの開発
- 3) Twitter のハッシュタグと位置情報の対応付け
- 4) 行為情報を地図上で可視化

下記にそれぞれの機能の実現手段、実現状況について述べる。

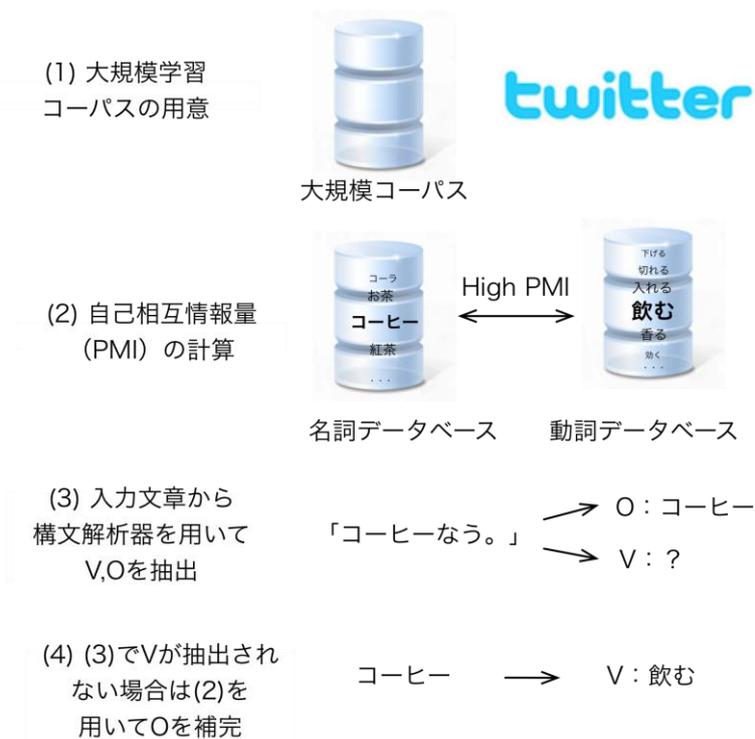


図 1: テキストからの行為情報抽出処理の流れ

### 1) Twitter からの行為情報の抽出

ウェブから Twitter が提供する API を通して、対象とする範囲の Tweet を収集し、自然言語処理を通じてテキストから動詞・目的語・主語のセットを一つの行為として抽出するエンジンの開発を行う。Twitter テキストから行為情報の抽出の流れを図に示す。



図 2: 建物内の行為を位置情報と共に取得するアプリケーションスクリーンショット

## 2) 建物内の行為情報を位置情報と共に取得するアプリケーションの開発

ユーザが携帯端末の GPS 機能を利用してテキストを投稿することにより、Twitter に投稿される情報からも位置情報を取得することは可能である。しかし、建物内の詳細な位置情報と共に、人々の行為を抽出することはできない。そこで、本プロジェクトでは、建物内においても、詳細な位置情報と共に Twitter に投稿するための専用アプリケーションの開発を行う。iPhone/iPod アプリケーションとして作成したソフトウェアのスクリーンショットを図 2 に示す。

## 3) Twitter のハッシュタグと位置情報の対応付け

2) で開発したアプリケーションによって投稿される Tweet は、ある特定の建物情報とハッシュタグを通じて関連付けられる仕組みになっている。この仕組みを使うことにより、GPS 付きの Tweet とシームレスに地図上に可視化することが可能となっている。

## 4) 行為情報を地図上で可視化

1) で収集し、抽出した行為データの動詞部分をタグ化し、Google Map API を用いて可視化している。テキストから動詞部分のみを可視化することにより、数多くの Tweet がある地点に投稿されても、煩雑になりすぎることなく地図上に表示することが可能となっている。タグクラウドとして Tweets 情報と共に表示されているシステムのスクリーンショットを図 3 に示す。

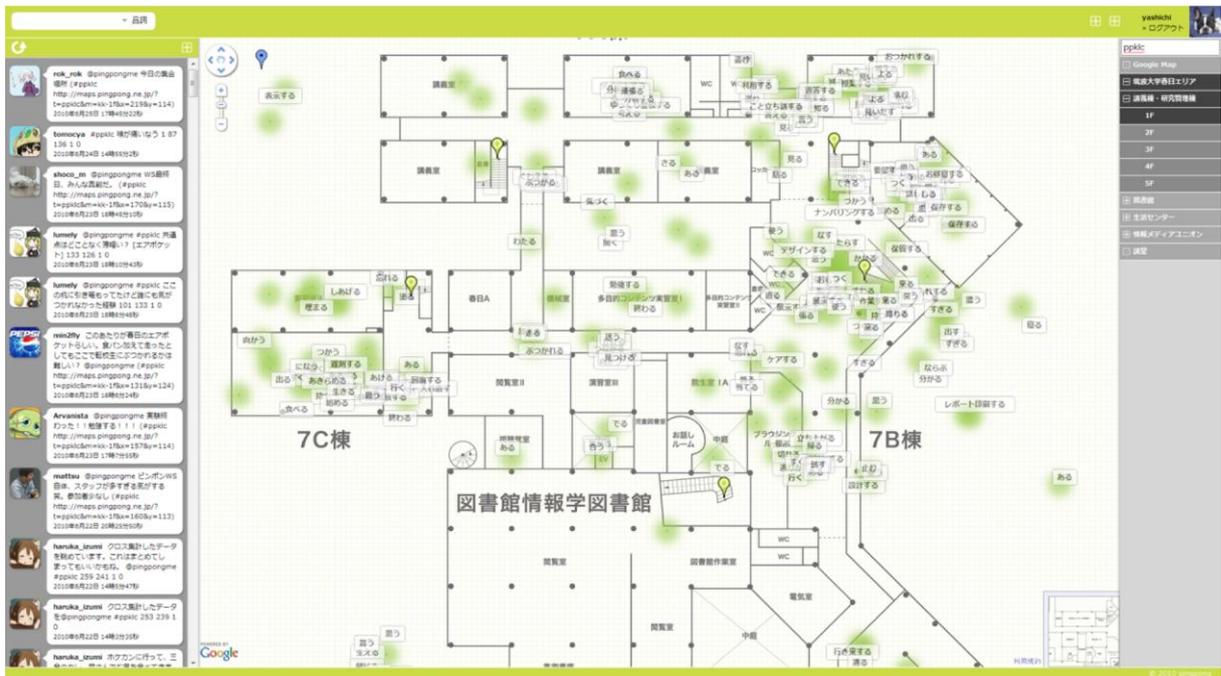


図 3 : 「動く地図」 タグクラウドと Tweets のタイムライン

## 10. 進捗概要

中間報告会でもコンセプトとプロトタイプを作成しており、ほぼ計画どおりプロジェクトは進捗した。当該プロジェクトは数多くの方が携わっているが、高いリーダーシップ力で、成果物を期間内に開発し完了させた。また期間後半には実証実験もでき、成果を振り返ることもでき、今後の展開に活かされる道筋ができた。

## 11. 成果

開発した「動く地図」を利用することにより、以下のような一般へのサービスと分野の創出が期待できる。

- (1) ショッピングモールなどの商業施設で本プロジェクトのシステムを導入することにより、例えば、非計画購買者(何を購入しようか、あらかじめ決めていない購買者)を動機づけ、購買行動を促進できる。
- (2) 「動く地図」の持つ時間的コンテキストの活用により、利用者の行為に反応したりリアルタイム性、即効性のある情報を入手できる。
- (3) 地図に紐づいた(tag つけられた)情報の取得・提供により、行ったことのない空間の情報も、高い精度で手に入れることができる。

## 12. プロジェクト評価

当該プロジェクトは人々が携帯端末で発信する情報をリアルタイムに取得し、地図と組み合わせることで「いま・ここ」で行われている行為を可視化する「動く地図」をつくるという挑戦的なテーマであった。実証実験でもある程度の評価がでて、難しいテーマで期間内にあるレベルまで完了させたことは大変評価できる。

実用の観点では幾つかの課題は残るが、可能性を感じるので、この技術をベースに今後も継続的に質を高め、研究だけではとどまらず、タイミングを考慮し、社会的に有用なレベルまでこの技術を完成することにチャレンジしてほしい。

また、人材育成的な観点においては本プロジェクトの経験をとおして、自己表現力、市場・ユーザ志向の考え方の方法、ビジネスマインド、創造力を向上できたと考えられる。新しいことにチャレンジする情熱もあり、研究開発者としてのポテンシャルも感じられるので、さらなる成長と活躍を期待している。

## 13. 今後の課題

開発した「動く地図」をウェブを通して公開していく。最終的には、ウェブ系の大企業と連携し、行為抽出エンジンを大規模に運用し、商業施設における空間レイアウト設計や、行動予測、実空間と結びついた広告サービスのアプリケーションとして提供することを目標としており、目標が達成されたときには多くのユーザに使われるシステムとして、情報産業への影響を与えることが可能であると考ええる。

しかし、さまざまなウェブ上でのサービスが日々立ち上げられている今日、多くのユーザの目を集めることは難しくなっており、実際にどのくらいのユーザに使われるかは実際にサービスを公開してからでないと分からない部分も多い。そこで、「動く地図」を、誰もが利用可能なサービスとして公開して発信・普及に努めると共に、大学や図書館といった公共空間への「動く地図」を導入するためのきっかけとして、ワークショップを活用していく予定である。今後も必要な時点で個別にアドバイスを行ってゆきたい。