

# ソーシャルメディアの情報統合による エキスパート検索エンジンに関する研究

## An expert retrieval by integrating social media

大澤 昇平<sup>\*1</sup>  
Shohei Ohsawa

松尾 豊<sup>\*1</sup>  
Yutaka Matsuo

<sup>\*1</sup> 東京大学

The University of Tokyo

An *expert* for a given topic is a person who has high level of skill for the topic. Recent grows of so-called crowd sourcing service requires the method to retrieve the highly-skilled expertise from a lot of candidate experts. An *expert retrieval problem* is a problem that finding the highly-skilled experts from given database in an organization. Several papers tackled this problem, however, the papers targets to at most a few handles experts as far. Practical expertise retrieval requires over ten thousand of experts. This paper proposes a framework for large-scale expert retrieval by integrating social media.

### 1. はじめに

近年、クラウドソーシングサービスやオープンソースシステム・コミュニティの発達により世界全体において人材流動性が高まってきており、特定のトピックの問題を解決することに対して高いスキルを有している人物、すなわち**エキスパート**からいかに専門性を推定するかということが仕事を発注する側の組織において焦点となっている。**エキスパート検索問題**[Balog *et al.*, 2006]は、組織の文書などを格納しているデータベースから、目的の専門知識を有したエキスパートを検索する問題である。エキスパート検索に関する分野は、fCHERを筆頭とするワークショップが開かれたり、評価用のデータセットがTRECから提供されたりするなど、盛んに研究が行われている。

従来より、エキスパート検索は行われているものの、一方で実証実験は数百人規模にとどまっており、クラウドソーシングサービスに応用するためには、より大規模な(数万人規模)のエキスパートに対しての実証実験を行うことが必要になってくると考えられる。

本研究の目的は、ウェブ上のデータソースを用いたエキスパートの検索基板、**Social Expert Search Engine** (以下、SESE)の実現である。SESEは、ウェブ上に分散するデータソースを収集し、統合することで、より広範な候補エキスパート集合に対して検索を可能にする。

### 2. 関連研究

エキスパート検索問題は、データベースから、目的の専門知識を有したエキスパートを検索する問題であり、この分野は主にBalogによって掃引されている。彼らの研究グループによる代表的な文献として、[Balog *et al.*, 2006]がある。この文献中において、Balogらは、文書集合からエキスパート検索を行うための、ふたつの一般的な方策について述べている。ひとつは、プロフィール中心アプローチと

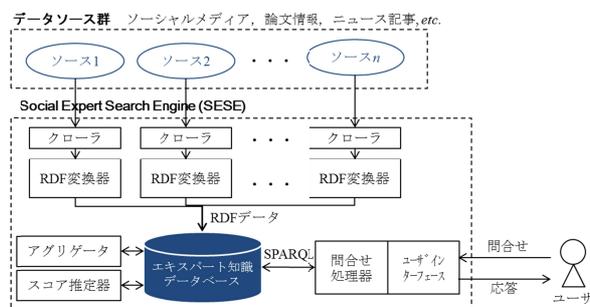


図1 提案システム(SESE)の全体像

呼ばれており、候補となるエキスパートを、文書に基づき直接的にモデルする、もうひとつは文書中心アプローチと呼ばれ、あるクエリにもとづき関連する文書を最初に検索し、次に関連するエキスパートを検索する。Campbellら[Campbell *et al.*, 2003]は、電子メールのやり取りによって構成されるネットワークに対して、グラフベースのエキスパート検索アルゴリズムを適用している。具体的には、HITSアルゴリズムを用いており、得られるオーソリティスコアによってエキスパートの専門性を推定している。Zhangら[Zhang and Li, 2010]は、研究者の共著ネットワークに対して、エキスパート検索を行っている。彼らは検索対象となる研究者集合のうち、少数の部分集合に対して教師データとなるスコアを与えた上で、信念伝搬モデルによって、そのスコアを共著ネットワーク上の他の研究者に対して伝搬させている。Karimzadehganら[Karimzadehgan *et al.*, 2009]は、組織におけるヒエラルキー(上司-部下関係および同僚関係によって構成される構造)に着目したエキスパート検索を試みている。得られた知見としては、各階層において近接性の高い研究者同士は、似たトピックに対して専門性を有しているということである。したがって、ただ単にソーシャルネットワーク上で専門性スコアを伝搬させるのではなく、周囲のユーザに対して伝搬させるのが良いという結果を得ている。これらのアプローチは、どれも数百人のネットワークを対象に実験を行っているが、本研究が提案す

連絡先: 大澤昇平、東京大学大学院工学系研究科、東京都文京区弥生 2-11-16 工学部 9 号館 210 ohsawa@weblab.tu-tokyo.ac.jp



図2 ポートフォリオ自動生成エンジン(PGE)の出力。一部スタブデータを用いている。

るSESEの枠組みでは、ウェブ上から自動的に収集したデータを用いることで、数万人規模のエキスパートに対して同手法を適用することが可能になる。

### 3. Social Expert Search Engine (SESE)

図1に、SESEの概要について述べる。クローラは、ウェブ上のデータソースにアクセスをし、取得した情報をRDF変換器に渡す。RDF変換器では、データソースの情報をRDF (Resource Description Framework)に変換し、エキスパート知識データベースに格納する。エキスパート知識データベースは、RDFとして情報を保持する。アグリゲータでは、エキスパート知識データベースに対し、レコードの統合を行う。スコア推定器では、エキスパートに対し専門性スコアの付与を行う。問合せ処理器とユーザ・インターフェースでは、ユーザから受けた検索要求を、RDFデータベースへの問合せに変換する。

SESEの応用例のひとつとして、**ポートフォリオ生成エンジン**(Portfolio Generation Engine; PGE)がある。PGEは、SESE上のデータベースに対して問合せを行い、出力を行う。著者らが開発したPGEの出力を、図2に示す。

### 4. 実験

本節では、PGEに対して評価を行う。評価は9名の被験者を対象に行う。各被験者は各自に対してPGEが生成したポートフォリオを閲覧したあと、12項目からなるアンケート項目に対して5段階評価(1-5の離散値)で答える。今回の実験においてSESEが対象とするウェブ上のデータソースは、GitHub (www.github.com), Google Play (play.google.com)である。本研究ではこれらのウェブサービスから独自にクロウリングを行っている。

表1 アンケート調査の結果。12項目の質問のうち、一部を抜粋している。

質問事項	スコア <sup>†</sup>
スキルに関する項目は、あなたのスキルに合っていましたか?	3.3±3.0
スキルに関する項目は、あなたのスキルを十分にカバーしたものでしたか?	3.0±2.8
一緒に働いた人の項目は、実際の内容を反映していましたか?	3.0±2.0
生成されたポートフォリオのデザインには満足していますか?	3.1±0.8
カバー画像(ポートフォリオの背景画像)ですが、種類を増やしたほうが良いでしょうか?	3.9±1.8
ポートフォリオ自動機能では、レポジトリの画像を自動的に取得して表示します。レポジトリの画像は、プロジェクトの内容に合っていましたか?	2.7±1.0
ポートフォリオを、自己紹介ページとして友人に見せたり、ソーシャルメディア(Twitter, Facebook, etc.)に貼り付けたいと思いますか?	2.1±1.4

<sup>†</sup>記号±以降の数字は、標準偏差の2倍を表す。

実験結果を表1に示す。これを見ると、スキルに関しては、全く合っていないと答えた人と、全く合っていると答えた人の間でばらつく結果となった。「全く合っていない」と回答したユーザは、ウェブ上にデータが少なく、スキルを正しく推定できていなかった。デザインに関しては概ね良いとも悪いとも言えない結果であった。ポートフォリオを他の友人に見せたいかどうかに関しては、低い結果となった。このようなコメントをした理由については、被験者に確認したところ、デザインを画一的なものではなくより自分自身を表現できるようにカスタマイズしたいということであった。

### 5. 結論と今後の課題

本研究では、エキスパート検索においてデータが少ないという問題を解決するため、ソーシャルメディアから自動的にデータを収集する枠組みSESEの構築を行い、その応用例としてPGEを構築し、9名の被験者に対して評価を行った。PGEはデータソースに掲載されている範囲でユーザの。今後の課題として、被験者から得られたコメントをSESEに対して反映する予定である。

### 参考文献

[Balog et al., 2006] K. Balog, L. Azzopardi and M. d. Rijke. Formal models for expert finding in enterprise corpora. In *Proc. SIGIR'06*, pp. 43-50, 2006.

[Campbell et al. 2003] C. S. Campbell, P. P. Maglio, A. Cozzi and B. Dom. Expertise identification using email communications. In *Proc. CIKM'03*, pp. 528-531, 2003.

[Karimzadehgan et al., 2009] M. Karimzadehgan, R. W. White, and M. Richardson. Enhancing expert finding using organizational hierarchies. In *Proc. ECIR'03*, pp. 177-188, 2009.

[Zhang and Li, 2010] J. Zhang, J. Tang, and J. Li. Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications*, Vol. 4443/2010, pp. 1066-1069, 2010.