

# Web コンテンツ中の社会問題と その解決ミッションの抽出手法の検討

Investigation of methods for extracting social issues and missions  
for solving them from Web content

神谷 晃<sup>1</sup> 白松 俊<sup>1</sup> 長谷川 徳賢<sup>1</sup>

Akira Kamiya<sup>1</sup>, Shun Shiramatsu<sup>1</sup>, Tokutaka Hasegawa<sup>1</sup>

<sup>1</sup>名古屋工業大学大学院 情報工学専攻

<sup>1</sup>Nagoya Institute of Technology Department of Computer Science

**Abstract:** In this paper, we propose methods for extracting social issues and missions from Web content like news articles. The purpose of this study is to solve the shortage of resources such as human resources of civil society organizations such as NPOs. Therefore, it is necessary for citizens to sympathize with the issues that NPO organizations want to solve. For this purpose, information extracted with reference to Golden Circle is defined, and extraction is automatically performed on the Internet by using machine learning. In this paper, we describe a technique for that purpose.

## 1. はじめに

人口減少・高齢化が進展する中、地域の社会課題に対応するためには、その地域の行政のみの取り組みだけでは限界があると考えられ、民間の人材・資金等のリソースを社会課題の解決のために取り込む必要性があると考えられる。そのためには、NPO 団体等の市民活動団体の活動が不可欠であるが、市民活動団体の代表である NPO 団体は、「人材の確保や教育」(66.9%)・「収入源の多様化」(54.2%)といった課題を抱えている [1]。市民活動団体のリソースを増加させるためには、その地域の市民と市民活動団体が協力する必要がある。しかし、市民が過去1年間のボランティア活動経験の割合が17.4%・過去1年間の寄附経験の割合が41.2%と、市民の社会貢献の割合は低い傾向にある[2]。これらは、市民活動団体に対する十分な情報がない事に起因している可能性が高い[2]。

この問題を解決するためには、地域課題や市民活動団体に関する情報を収集し、適切に表示するシステムが要求されると考えられる。われわれはこれまで、課題解決に関する活動をしている本人が活動内容(ミッションや解決アプローチ)を入力できる MissionForest という Web アプリケーションを開発してきた[5,6]。しかし、本人が入力するだけではスケーラビリティに欠け、広域的な地域課題解決には繋がりにくい。そこで本研究では、これらの情報を Web 上から自動で抽出することを目指す。本稿では特に、Web 記事にアノテーションすることで訓練コーパスの構築を目指す。

## 2. 事前知識

### 2.1 シビックテック

シビックテックとは、シビック(Civic:市民)とテック(Tech:テクノロジー)をかけた造語である。その地域に住む市民自身が、テクノロジーを活用し、行政サービスの問題や社会課題を解決しようとする取り組みのことである。本研究の最終的な目標は、こういった活動の支援を目的とする。

### 2.2 Golden Circle

課題を解決しようという活動において、協力者を募るために必要な情報は何だろうか。本研究では社会課題やその解決のためのミッションを Web から収集するが、どんな情報が必要かを考える上で参考となるのが、Sinek の Golden Circle[3,4]である。

Golden Circle は、中心から順番に why, how, what の3つの円から構成されるフレームワークである。このフレームワークは、人は「何を(What)」ではなく「なぜ(Why)」に動かされるので、Why → How → What の順で考え、その順番で説明を行うと共感を得られるといったことを表したものである。市民活動団体が解決したい課題に対し、仲間集めを行う際には、課題に対し共感してもらう必要があると考えられる。本研究では、この Golden Circle に基づき、ニュース記事等から情報の抽出を行う。我々は、これに加え where と who が必要であると考え。なぜならば、「誰がやったのか」・「どこでやったのか」が周知されることにより、様々な市民活動団体の共創につながると考えられるからである。

## 2.3 MissionForest

MissionForest [5,6]は、複数人での協働支援を目的とした Web アプリケーションである。複数人で共通の課題を解決するために課題を達成するためのタスクを Work Breakdown Structure (WBS) の形式で共同編集を行う (図 2)。本研究では、抽出をした社会課題をこのアプリケーション上で表示することを最終的な目的とする。

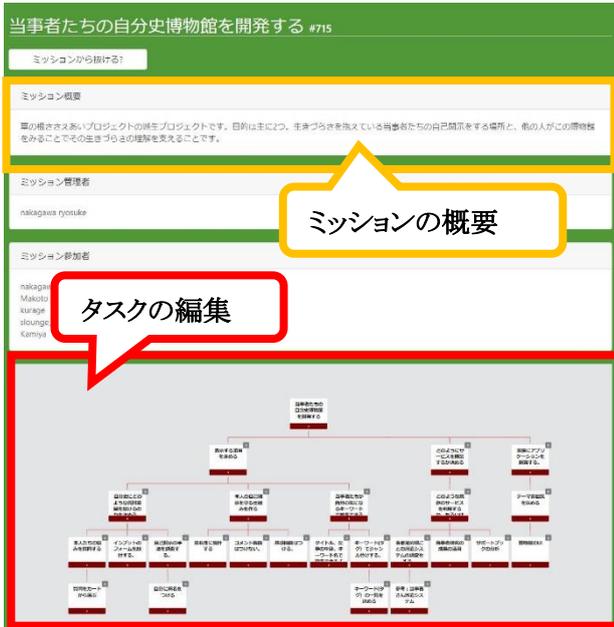


図 2 MissionForest

## 3. 提案手法

### 3.1 抽出するデータの定義

市民活動団体に協力してもらうためには、その団体が掲げる課題に対し、共感してもらう必要がある。そのために、インターネット上の地域課題の記事に関して Golden Circle を基に以下の情報を抽出すると良いと考えた。

- 1 Why (なぜ行うのか)
  - 1.1 どのような地域課題か?
  - 1.2 最終的にどのような社会にしていきたいのか?
  - 1.3 その地域課題のもたらすデメリットは?
  - 1.4 どの地域の地域課題か? (where)
  - 1.5 その地域課題を解決すると誰が嬉しいのか?
- 2 How (どのように)
  - 2.1 どのような方法で解決しようとしているのか? (解決アプローチ)
- 3 What (何を)
  - 3.1 実際に団体が何をしているのか?
- 4 Who (誰が)
  - 4.1 市民活動団体の名前

上記に基づき、doccano というツールを用い記事をアノテーションすると図 3 のようになった。

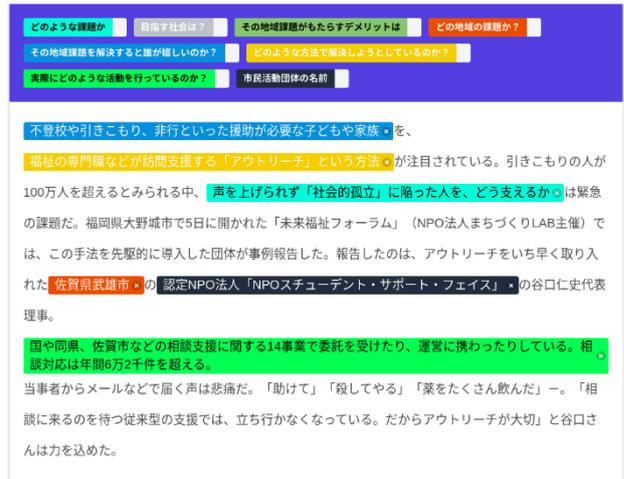


図 3 アノテーション例

### 3.2 社会課題データの公開

我々は、Web 上にある地域課題に関する機械・人手に関係なくアノテーションを行ったデータを、Web Annotation Data Model[6]の形式で公開することで、情報抽出の研究コミュニティで、そのデータを流通させることが可能であると考えられる。Web Annotation Data Model に従うと以下の形式 (図 4) でアノテーションデータの公開が可能であると考えられる。以下のアノテーションされたテキスト (対象テキスト) が、何のインスタンスかを表現したい。以下の例の場合、wikidata の引きこもりとリンクすると考えられる。

```

/* 省略 */
"selector": [
  {
    "type": "RangeSelector",
    "endOffset": 29,
    "startOffset": 1,
    "endContainer":
      "/div[1]/div[3]/main[1]/article[1]/div[2]/div[1]/div[2]/p[2]",
    "startContainer":
      "/div[1]/div[3]/main[1]/article[1]/div[2]/div[1]/div[2]/p[2]"
  },
  {
    "end": 4654,
    "type": "TextPositionSelector",
    "start": 4671
  },
  {
    "type": "TextQuoteSelector",
    "exact": "声を上げられず「社会的孤立」に陥った人を、どう支えるか?",
    "prefix": "100万人を捉えると超えるとみられる中",
    "suffix": "は緊急の課題だ"
  }
]
/* 省略 */

```

図 4 Web Annotation Data Model の一例

そのため、上記のアノテーションの URI が <http://srmlab.org/anno1> であった場合、上記のデータは以下のように表現することが可能である。

```

{
  "@context": "http://www.w3.org/ns/anno.jsonld",
  "id": "http://srmtlab.org/anno1",
  "type": "Annotation",
  "body": "http://www.wikidata.org/entity/Q224156",
  "target":
    "https://www.nishinippon.co.jp/item/n/551563/",
  "motivation": "classifying"
}

```

図5 図4の対象テキストとwikidataをリンクさせた例

### 3.3 アノテーションツールの作成

現在、Web上のデータをWeb Annotation Modelに従いアノテーションを行うツールとして、hypothes.isと呼ばれるツールが存在する。これは、Hypothesisと呼ばれる非営利のコミュニティが行っているプロジェクトで、特別な実装なしで全てのWebページでアノテーションを行えるようにするプロジェクトである。このコミュニティは現在、chromeの拡張機能を用いたアノテーションツールをソースコードと共に公開している。我々は、このツールを拡張し、wikidataとリンクを行えるような拡張機能を開発・追加し、アノテーションデータを拡張するつもりである。

### 3.4 アノテーション

我々は、Web上のデータにアノテーションを行うために、12月11日にアノテーションデータソンを開催する予定である。そして、その後Yans2019で行われたアノテーションハッカソン[7]を参考に、我々もアノテーションをハッカソン形式にし、質の良いアノテーションデータを取得したいと考えている。

### 3.5 課題可視化機構の設計

- 本研究の最終的な目標は、設定した課題を収集した地域課題に関する情報をMissionForestに表示することにある。収集した情報を以下のルールに基づき表示を行いたい。
- ミッションの概要と、一番上位のタスクに抽出した情報のWhyの部分について書く
- 第2階層目(図6の緑色)のタスクに抽出した情報のHowの部分について書く
- 第3階層目以下の階層(図6の赤色)のタスクに直近で何をやるか・何をやってきたのかを書く

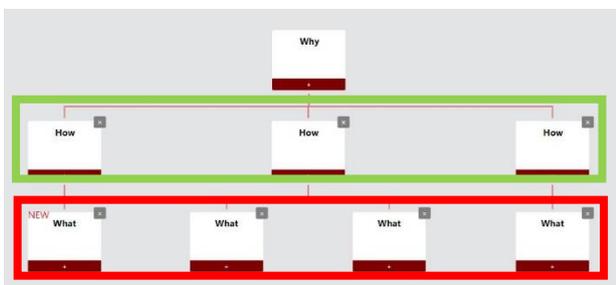


図6 MissionForestの階層構造

## 4. 最後に

今回、インターネット上にあるニュース記事等から地域の社会課題に関する情報を抽出する手法について提案した。今後は、インターネット上にある記事にアノテーションを行い、そしてアノテーションしたデータに、自然言語処理の技術を用いることで学習モデルを作成する予定である。

## 謝辞

本研究の一部は、JSPS 科研費(17K00461)の支援を受けた。

## 参考文献

- 内閣府 政策統括官. “平成 29 年度 特定非営利活動法人に関する実態調査”,平成 30 年 3 月 30 日公表
- 内閣府 政策統括官. “平成 28 年度 市民の社会貢献に関する実態調査”,平成 29 年 3 月 21 日公表
- Sinek, S. “How great leaders inspire action”, [https://www.ted.com/talks/simon\\_sinek\\_how\\_great\\_leaders\\_inspire\\_action](https://www.ted.com/talks/simon_sinek_how_great_leaders_inspire_action) 2019 年 7 月 3 日アクセス。
- サイモン・シネック. “FIND YOUR WAY あなたとチームを強くするシンプルな方法” ディスカバー・トゥエンティワン, 2019
- 後藤誉昌, 渡辺賢, 白松俊. “MissionForest: 組織内外における協働支援のためのタスク構造化システムの試作”, 第 79 回情報処理学会全国大会, 6ZA-02, 名古屋大学東山キャンパス, 2017.3.
- 渡辺賢, 白松俊, 仙石晃久. “共創支援システム「みらいらぼ」とタスク構造化システム「MissionForest」の相互連携による共創継続支援の検討”, 人工知能学会 第 3 回市民共創知研究会(SIG-CCI) 「みらいらぼ なごや」, 名古屋市 西祐寺, 2017.12.
- W3C. “Web Annotation Data Model”, <https://www.w3.org/TR/annotation-model/> 2019 年 11 月 11 日アクセス
- @ymym3412. “YANS2019 でアノテーションハッカソンを開催しました”, <https://note.mu/xpaperchallenge/n/n7782502267db> 2019 年 11 月 8 日アクセス