

(2)「現行の放射性物質に関する食品出荷基準は、諸外国に比べて緩（ゆる）い」

※(2)の間ずっとカード⑤提示

Q15. 最初にこの情報をどこから聞きましたか。(選択肢読み上げ：ひとつだけ)

- |                   |                       |
|-------------------|-----------------------|
| 1. TV・新聞などのマスコミ発表 | 5. その他(具体的に: )        |
| 2. インターネットの情報     | 6. 覚えていない             |
| 3. 家族・友人などからの伝聞   | 7. 聞いたことがない (⇒p5(3)へ) |
| 4. 公的機関の広報        |                       |

Q16. 最初にこの情報を聞いたとき、この情報は間違っただけの情報と聞きましたか、正しい情報と聞きましたか。(選択肢読み上げ：ひとつだけ)

- |              |            |
|--------------|------------|
| 1. 正しい情報     | 3. わからなかった |
| 2. 間違っただけの情報 |            |

Q17. 最初にこの情報を聞いたとき、あなたは正しい情報だと思いましたが、それとも誤った情報だと思いましたが。(選択肢読み上げ：ひとつだけ)

- |                   |            |
|-------------------|------------|
| 1. 正しい情報だと思った     | 3. わからなかった |
| 2. 間違っただけの情報だと思った |            |

Q18. この情報について、いまもそう思っていますか。(ひとつだけ)

- |                 |                    |
|-----------------|--------------------|
| 1. はい (⇒p5(3)へ) | 3. わからなくなった (⇒SQへ) |
| 2. いいえ (⇒SQへ)   |                    |

**SQ.【Q18で「いいえ」「わからなくなった」とお答えの方に】**

何が原因で判断が変わりましたか。あてはまるものをいくつでもお答え下さい。

(選択肢読み上げ：いくつでも)

- |                   |                |
|-------------------|----------------|
| 1. TV・新聞などのマスコミ発表 | 4. 公的機関の広報     |
| 2. インターネットの情報     | 5. その他(具体的に: ) |
| 3. 家族・友人などからの伝聞   | 6. 覚えていない      |

(3) 「一般に販売されているガイガーカウンター（放射線を計測する機器）を一食分の食品にそのままあてても、基準を超えたかどうか判定できない」

※(3)の間ずっとカード⑥提示

Q19. 最初にこの情報をどこから聞きましたか。(選択肢読み上げ：ひとつだけ)

- |                   |                       |
|-------------------|-----------------------|
| 1. TV・新聞などのマスコミ発表 | 5. その他(具体的に: )        |
| 2. インターネットの情報     | 6. 覚えていない             |
| 3. 家族・友人などからの伝聞   | 7. 聞いたことがない (⇒p6(4)へ) |
| 4. 公的機関の広報        |                       |

Q20. 最初にこの情報を聞いたとき、この情報は間違っただけの情報と聞きましたか、正しい情報と聞きましたか。(選択肢読み上げ：ひとつだけ)

- |              |            |
|--------------|------------|
| 1. 正しい情報     | 3. わからなかった |
| 2. 間違っただけの情報 |            |

Q21. 最初にこの情報を聞いたとき、あなたは正しい情報だと思いましたが、それとも誤った情報だと思いましたが。(選択肢読み上げ：ひとつだけ)

- |                   |            |
|-------------------|------------|
| 1. 正しい情報だと思った     | 3. わからなかった |
| 2. 間違っただけの情報だと思った |            |

Q22. この情報について、いまもそう思っていますか。(ひとつだけ)

- |                 |                    |
|-----------------|--------------------|
| 1. はい (⇒p6(4)へ) | 3. わからなくなった (⇒SQへ) |
| 2. いいえ (⇒SQへ)   |                    |

**SQ.【Q22で「いいえ」「わからなくなった」とお答えの方に】**

何が原因で判断が変わりましたか。あてはまるものをいくつでもお答え下さい。

(選択肢読み上げ：いくつでも)

- |                   |                |
|-------------------|----------------|
| 1. TV・新聞などのマスコミ発表 | 4. 公的機関の広報     |
| 2. インターネットの情報     | 5. その他(具体的に: ) |
| 3. 家族・友人などからの伝聞   | 6. 覚えていない      |

(4)「原発事故後、新たに食品に加えられた放射性物質より、元から入っている放射性物質のほうが多い」

※(4)の間ずっとカード⑦提示

Q23. 最初にこの情報をどこから聞きましたか。(選択肢読み上げ：ひとつだけ)

- |                   |                        |
|-------------------|------------------------|
| 1. TV・新聞などのマスコミ発表 | 5. その他(具体的に: )         |
| 2. インターネットの情報     | 6. 覚えていない              |
| 3. 家族・友人などからの伝聞   | 7. 聞いたことがない (⇒p7 Q27へ) |
| 4. 公的機関の広報        |                        |

Q24. 最初にこの情報を聞いたとき、この情報は間違っただけの情報と聞きましたか、正しい情報と聞きましたか。(選択肢読み上げ：ひとつだけ)

- |              |            |
|--------------|------------|
| 1. 正しい情報     | 3. わからなかった |
| 2. 間違っただけの情報 |            |

Q25. 最初にこの情報を聞いたとき、あなたは正しい情報だと思いましたか、それとも誤った情報だと思いましたか。(選択肢読み上げ：ひとつだけ)

- |                   |            |
|-------------------|------------|
| 1. 正しい情報だと思った     | 3. わからなかった |
| 2. 間違っただけの情報だと思った |            |

Q26. この情報について、いまもそう思っていますか。(ひとつだけ)

- |                  |                    |
|------------------|--------------------|
| 1. はい (⇒p7 Q27へ) | 3. わからなくなった (⇒SQへ) |
| 2. いいえ (⇒SQへ)    |                    |

SQ.【Q26で「いいえ」「わからなくなった」とお答えの方に】

何が原因で判断が変わりましたか。あてはまるものをいくつでもお答え下さい。

(選択肢読み上げ：いくつでも)

- |                   |                |
|-------------------|----------------|
| 1. TV・新聞などのマスコミ発表 | 4. 公的機関の広報     |
| 2. インターネットの情報     | 5. その他(具体的に: ) |
| 3. 家族・友人などからの伝聞   | 6. 覚えていない      |

●次に放射能に関するご自身の意見や立場についてお伺いします。

Q27. 食品の放射能汚染について、お伺いします。

a) 食品の安全性について、どのような情報を求めていますか。(自由回答)

b) その情報は、国が提供すべきだと思いますか、思いませんか。どうしてそのように思うかもあわせて、ご意見をお聞かせ下さい。(自由回答)

<国が提供すべきかどうか>

<どうしてそのように思うか>

c) 国が、どのような情報を、どんな形で提供してくれればいいと思いますか。  
(自由回答)

Q28. 仮に食品の買い物の際欲しかった品物に「福島産」と書かれていたら、どうしますか。

**(自由回答)**

※ 「食品による」という回答がでた場合、重ねて質問

Q.どのような食品なら「福島産」と書かれていても購入しますか。

Q29. あなたは、原子力エネルギーの利用は今後も進めていく必要があると思いますか。

**(選択肢読み上げ：ひとつだけ)**

- |              |               |
|--------------|---------------|
| 1. とてもそう思う   | 4. あまりそう思わない  |
| 2. ある程度そう思う  | 5. まったくそう思わない |
| 3. どちらともいえない |               |

Q30. 放射能は、あなたの暮らしにどのように影響を与える、または与えていると思いますか。自由に考えをお聞かせ下さい。**(自由回答)**

～ 調査は以上で終了です。ご協力、ありがとうございました。

## インターネットにおける意見解析手法の信頼性に関する研究

研究分担者 藤井 仁

（国立保健医療科学院研究情報支援研究センター 主任研究官）

研究分担者 奥村 貴史

（国立保健医療科学院研究情報支援研究センター 特命上席主任研究官）

### 研究要旨

ネット情報の解析による意見分析は、一般的に、定量的には意義を見出すことが困難である。しかしながら、ネットの利用者は増大を続けており、マスコミ情報から口コミ情報まであらゆる情報がネット上に表出されていることから、ネットにおける意見の定性的な分析は、少ないサンプルの社会調査以上に網羅性を有するのではないかと考えられる。そこで本研究分担では、インターネットにおける意見解析手法の信頼性を検証するため、ネット情報の解析により得られる意見と実社会に見られる意見との関係性について検討を試みた。まず、地方都市や高齢者、専業主婦を対象とした対面方式のアンケート調査を実施し、放射能汚染に関する 1,050 件の意見を聴取した。その上で、アンケートにより取得した放射能に関するそれぞれの意見について、ネット上に存在するかの検証を行った。その結果、インタビュー調査により聴取した意見のほとんどに、インターネット上に対応する意見を見出すことが出来た。限られた人数を対象に行った今回の予備調査においては、方法論上の限界はあるものの、ネット情報の解析は社会を知るうえである程度信頼に足る手段であることが示唆された。今後、ネット情報に対する自然言語処理をリスクコミュニケーションの実務へと役立てる実用化研究が望まれる。

### A. 研究目的

2011 年 3 月に生じた東京電力福島第一原発事故により、環境中に大量の放射性物質が放出された。こうした前例の無い規模の放射能災害の発生に際して、国や東京電力を初めとする公的機関による情報提供上の問題もあり、発災直後から現在に至るまで、放射能に関する正誤のあいまいな情報や伝聞情報がインターネットを中心として蔓延する事態が生じている。

このように健康への大規模な危険が生

じた際の公的機関による情報提供は、リスクコミュニケーションとして知られ、今までにも、環境汚染による風評被害などに対応するために検討され、実践されてきた。しかしながら、従来のリスクコミュニケーションは、マスメディアを通じた一方通行な情報伝達、ないし、専門家と一般人の双方向な交流のいずれかを前提としており、近年のようにネットを介して、公的機関や専門家、一般人が入り乱れた情報発信や消費をするような状況を想定していない。そのため、既存のリスクコミュニケーション

が成り立たなくなっている可能性があるにもかかわらず、ネット上に拡散している情報は膨大であり、統計学に基づく従来の社会調査手法では事態の全体像を把握することすら困難である。

そこで、ネット時代のリスクコミュニケーションを再構築していく上で、多量の自由記載文書を効果的に処理しうる手段である自然言語処理技術が期待されることになる。実際、自然言語処理を応用したネットの解析技術は、企業によるマーケティング目的等で既に商用利用されており、本研究班においても同様の技術を利用したネット上のさまざまな意見の解析を進めている。しかしながら、Twitter や Facebook などに表出されるネット上の意見と、実社会の世論調査結果には、大きな乖離があることは繰り返し指摘されている。ネット上の意見の多くは、意図的な成りすましや意見の水増しだけでなく、意見が自由に複製されうることから、とある意見が多いか少ないかといった定量的な判断に用いることが困難である。また、そもそもネットの利用は PC やスマートフォンなどの情報機器を利用できる年齢層に偏っており、表出されている意見の代表性については大いに疑念がある。さらに、調査により意見を受動的に抽出される社会調査と比べて、ネット上では能動的な意見表明がなされることから、後者はより態度が先鋭化する傾向が懸念される。

そこで、本研究分担は、実社会を対象とした意見調査とインターネット側での調査との比較対照を通じ、インターネットにおける意見解析手法の信頼性についての検討を行うことを目的とする。

## B. 研究方法

ネット情報の解析による意見分析は、上述の理由により、定量的には一般的に意義を見出すことが困難である。一方で、放射能汚染に関して現在なされている言論の全体像を知る上で、もはやネットの意見は

無視しえない。また、テレビやラジオなど、マスメディアにおける言説のほとんどは直ちにネット上にコピーされ、それに対する賛否双方からのさまざまな議論が形成されることから、現代では、実社会とネットの言論はほぼ連続していると考えることが出来る。さらに、ネットを利用した調査においては、多少の規模での社会調査を行う場合よりもはるかに多様な意見を効率的に収集することが可能である。したがって、ネットにおける意見の解析は、定性的には、社会調査以上に社会を代表し、また、意見間の関係を明らかにする網羅性を有するのではないかと考えられる。

とりわけ、実社会における意見がすべてネット上に存在するのであれば、ネット上の意見を定性的に分析することにより、社会における放射線に関する意見を効率的に分析することが可能となる。そこで本研究分担では、社会調査によって得られる実社会における意見のほとんどはネット上にも存在することを確認し、ネット調査の定性的な妥当性について検証を試みた。

この「実社会には存在するもののネット上には存在しない意見」の存在を検証するうえでは、ネットを日常的に利用している層ではなく、出来る限りネットへの接触が薄い層を対象とした調査とすることが効果的である。そこで、調査においては、地方都市や高齢者、専業主婦をターゲットとした対面方式のアンケート調査を企画した。そのうえで、アンケートにより収集した放射能に関する意見について、それぞれがネット上に存在するか逐次検証を行った。

意見の抽出に際しては、具体的なトピックを提示し賛成や反対などの態度を聴取する形式の調査が考えられるが、このような構造化された調査は、集計における効率化が図れる一方で、独自性の高い意見を拾い損ねる懸念がある。しかしながら、自由記載を中心とした非構造的な調査では、集計のためのコストが増し、客観性が損なわれる可能性がある。このように、両手法

Q27. 食品の放射能汚染について、お伺いします。

- a) 食品の安全性について、どのような情報を求めていますか。
- b) その情報は、国が提供すべきだと思いますか、思いませんか。どうしてそのように思うかもあわせて、ご意見をお聞かせ下さい。
- c) 国が、どのような情報を、どんな形で提供してくれればいいと思いますか。

Q28. 仮に食品の買い物の際欲しかった品物に「福島産」と書かれていたら、どうしますか。

Q30. 放射能は、あなたの暮らしにどのように影響を与える、または与えていると思いますか。自由に考えをお聞かせ下さい。

図 1 放射能汚染に関するアンケート

には一長一短があるが、今年度は、予備調査として、学問的な正統性やマーケティング的な妥当性よりは、今後の調査を設計するための大まかな傾向の把握を目的とし、あえて非構造化インタビューにより意見を聴取する形をとった。図 1 に、調査に用いた質問票を示す。

(倫理面への配慮)

本アンケート調査は、「震災後のデマ情報の情報源と、その拡散、終息に寄与する要因の分析」として国立保健医療科学院倫理審査委員会より承認されている調査(NIPH-IBRA#12037)の一部である。

### C. 研究結果

本研究は、本研究班の分担研究として設定された震災後に生じた放射能に関するデマ情報についてのアンケート調査結果を解析する形で行われた。アンケート調査は、東日本大震災、並びに、放射能災害の影響が少なかった地域から、大都市部と地方都市の代表として、東京、秋田、高知を選挙して実施され、それぞれ 70 件の回答を得た。アンケートに際しては、調査地域において、5 世帯おきに訪問し調査への協力を要請するランダムウォーク法を用いた。

調査対象者は、20代 4.8%、30代 12.9%、40代 16.2%、50代 18.6%、60代以上 47.6%、平均年齢は 56.6 歳と、年齢分布がかなり高齢者へと偏っている。また、男性 36.2%、女性 63.8%と、回答者の 2/3 が女性であった。学歴で見ると、中学卒 13.3%、高校・専門学校・短大卒 65.7%、大学・大学院卒 20.5%と、8 割が高等教育を受けていない。インターネットの利用に関しては、毎日利用していると答えたのが 31.4%で、利用しないと答えたのが 46.7%と半数近くを占めた。その他、週 1 回以上が 12.9%、週 1 回未満が 9.0%と、インターネットを日常的には利用していない者の割合が 7 割近くを占めており、研究の目的通りインターネットの利用が少ない層を抽出することが出来た。

本研究では、以上の対象者より、「食品の安全性に関して欲しい情報」、「その情報を国が提供すべきかどうか、また、なぜそう考えるか」、「国に、どのような情報を、どのような形で提供して欲しいか」、「欲しかった品物が福島産であった場合の対応」、「放射能が暮らしに与えている影響」の 5 つの質問について自由な意見を求めた。園上で、集まった 210 件の意見を目視にて類似意見にクラスタ化したうえで、クラスタ中の代表的な意見について、検索エンジン



意見	件数
食品には、安全だとわかるようにしっかりとした基準値を表示してほしい	52
正確な数字を隠蔽することなく開示してほしい	40
市場に出回っているものは安全だと思っているので、気にしていない	38
産地をはっきりと表示し、安全かどうかわかるようにしてほしい	26
健康にどのような悪影響が出るのか	14
すべての食品を検査して、すべてに放射能測定値を表示してほしい	9
騒ぎすぎ、情報に振り回されている	8
よくわからないのでテレビなどでわかりやすく教えて欲しい（具体的にどのような情報かはわからない）	6
もう歳なので、気にしない	4
グルーピングできない意見	4
分析対象外（意見になっていない／質問の意図とずれているなど）	8
Q27bの意見の混入	1

表 1. 食品の安全性について、どのような情報を求めているか (Q27a)

を用いて類似意見の探索を行った。また、クラスタに分類しえない意見については、それぞれ同様の方法で類似意見の探索を行った。最後に、より客観性を高めるため、別の調査者により意見の類似性の判定を行った。

5つの質問に対する各210件の回答をクラスタ化した結果を表1～5に示す。回答は10～15個のクラスタに分類され、5～10%ほどが分類不能な意見であった。作業においては、技術的、時間的な制約から、検索数を削減し、大幅な簡略を行ったが、検証した範囲においては社会調査による抽出意見のほとんどに対して、ネット上の類似意見を見出すことが出来た。より厳密な判定や他分野での意見により別の結果が得られる可能性はあるものの、大まかな傾向としては、ネット調査により抽出される意見の多様性は、社会調査により抽出される意見を上回るものと考えられる。検証結果の詳細は、別途、調査報告書に詳述されて

意見	件数
国に責任、義務がある	65
国が提供するのが安心	43
利害関係なく、公正かつ信頼できる情報を出せる	23
個人、民間では限界があり、国でないと対応できない	19
国が信頼できない	6
国の方が基準を統一しやすい	5
国がやるには範囲が広すぎるので、県や地域が対応するのがよい	5
生産者や販売店が責任を持って提供してほしい	5
税金の無駄使いになる	3
国が県や企業、生産者を指導して義務付けて欲しい	3
グルーピングできない意見	1
分析対象外（意見になっていない／質問の意図とずれているなど）	21
Q27aの意見の混入	8
無回答	3

表 2. 放射能の情報を国が提供すべきか？ どうしてそのように思うのか (Q27b)

いる。

#### D. 考察

今回、限られた人数を対象に行った予備調査として、類似意見を手作業により検索、確認する大まかな検証を行った範囲では、実社会に存在する意見のほとんどは、ネット上に対応する意見を見出しうる明らかとなった。歴史的に見ると、インターネットが研究教育用ネットワークであった80年代から、社会への普及が始まった90年代、ネットに接続可能な携帯電話の普及も合わせり社会に浸透した2000年代を経て、2010年代では、地デジ放送やタブレット端末等の普及により特に意識をしなくても誰もがネットに接続するようになりつつある。この傾向は今後も続くと考えられるため、実社会において発出される意見の多くは、ますますネット上でも観察される機会が増えていくと考えるのが妥当であろう。そうで

意見	件数
メディア、マスコミ、テレビや新聞	68
テレビ	28
自治体、自治体の広報誌	9
食品に表示	7
ネットやマスコミ、ネットやテレビ	6
新聞	5
インターネット	5
店頭	4
あらゆる手段、いろいろな手段	4
NHK	2
分析対象外（意見になっていない等）	26
無回答	46

表 3. 国が、どのような情報を、どのような形で提供すれば良いと思うか (Q27c)

ない意見は、地域性が極めて高いか、一定の集団に閉じた意見、ないし、私的な内容を含むなどの特殊な事例に限られ、絶対数として少ないだけでなく、そもそも社会調査により抽出すること自体が困難な意見である可能性が高い。

ただし、本手法を用いて確定的な結論を導くには、方法論上の問題が少なくない。まず、そもそも類似意見とは何か、という本質的な問題がある。「とある公園にホットスポットがあり心配」というアンケート結果が得られた際、ネットに完全な同一意見が見当たらなくても、「〇〇市にホットスポットがある」といった記述や「ホットスポットが心配」といった記述が認められれば、本研究の趣旨には合致するが、意見の抽象度や類似度の定義には恣意性が伴う。

また、質問を放射能汚染に関する5種類の質問に限ったが、別のテーマについて質問することで、得られる結果が変わる可能性は十分ある。さらに、今年度の研究では、コスト的な観点から、類似意見検索の効率化のために意見のクラスタ化を行ったうえで検索操作を行ったが、この操作には客観性と再現性に難がある。方法論的には、同一タスクを複数名で行い、意見の一致度を

意見	件数
安全だという検査情報が表示されていれば買う	32
福島産だろうが、売っている以上、安全なはず	22
福島産かは気にしない	20
地元産や産直を優先して買っている。自家栽培のものを利用	15
なんとなく躊躇する・できれば買いたくない	15
子どもには心配だが大人は大丈夫	12
生鮮食品(水産物、野菜など)は心配なので買わない	11
はじめは躊躇したが、いま(今後は)買う	10
安全情報がはっきりしていないから買わない	9
福島産以外の他県のを優先して買う	6
復興支援というかたちなら購入したい・福島の関係者なので応援したい	5
中国産など輸入品のほうが心配	3
放射能が心配	3
食品への放射能の蓄積が心配	2
分類不能	4
意見とは考えられない内容	41

表 4. 欲しい食品に福島産と書かれていた際、どうするか (Q28)

確認する等の操作が不可欠であろう。この問題は、意見の同一性判定においても当てはまる。客観性の担保のためには、利害関係のない第三者により判定を行うプロトコルとするか、類似意見検索を機械的に行うための処理系を開発するといった工夫が求められるだろう。このように、方法論的には、本手法は予備研究の域を出ない。

しかしながら、ネット調査により収集する意見には十分な網羅性が備わっていると看做すことが出来れば、新たなリスクコミュニケーション手法が実現する可能性が生じる。たとえば、本研究班の他研究分担で研究を進めているネット調査により抽出した意見間の関係を明らかとする「言論マップ」技術によって、放射線に関する誤った情報の根拠や意見間の対立関係等を可視化し、訂正情報の効果的な公表や対立解消に向けた働きかけ等が可能となる可能性がある

意見	件数
特にない・影響ない・よくわからない	40
原発についての賛成意見、反対意見	32
子どもや孫の世代や将来の蓄積が不安だ	25
放射能は目に見えないから怖い。わかりにくい	20
健康への不安がある	17
身近でないので、あまり考えていない。わからない	17
放射能は医療面で役に立っている	10
安全基準をしっかりと決めてほしい。情報を公開してほしい	10
放射能は危険	10
電気料金の値上がりが心配	6
食品や水への影響が心配	6
しかたない	3
産地を気にするようになった	2
放射能の瓦礫問題が不安	2
風評被害が心配	2
震災の影響が大きい。取り返しのつかない問題	2
甲状腺ガンと関係がある	2
単なる事実確認	1
その他	3

表 5. 放射能はあなたの暮らしにどのような影響を与えているか？ (Q30)

る。あるいは、誤った情報が次々と生成されているような状況下では、従来のリスクコミュニケーションでは対応できなかった問題に、速報性に優れた意見分析技術が有用となる可能性がある。

## E. 結論

本研究分担は、ネット時代のリスクコミュニケーション手法を確立する一助とするため、ネット情報の解析により得られる意見と実社会に見られる意見との関係性について検討を試みた。その結果、方法論上の限界はあるものの、大まかな傾向としてネットの定性的分析は実社会における意見を包含しており、社会を知るうえで程度信頼に足る手段であることを明らかとした。今後の方向性としては、この暫定的な結論における方法論上の課題を克服し仮説のよ

り精緻な検証を目指す方向と、ネットの定性的分析は社会を知るうえで信頼に足るといふ仮定に立ち、ネットにおいて表明される意見を実際のリスクコミュニケーションに役立てる方向がある。

前者としては、まず、質問のテーマや数、意見聴取を行うサンプル数を増やす必要がある。また、類似意見検索における客観性、再現性を高める手法の開発を行う必要がある。たとえば、既存の検索エンジンを用いて類似意見を検索するには、意見を適切な検索用キーワードに変換したうえで検索する必要がある。また、既存の検索エンジンを用いる限り、そのキーワードに関連した人々の意見ではなく、そのキーワードに関する「有用な情報」ばかりが得られることになる。これらの問題は、共同研究機関である NICT の WISDOM システムを用いることで、より効率化を図れる可能性がある。また、自然言語処理技術を活用することにより、類似意見検索の精緻化や意見類似度の定量的評価に取り組む道も考えられる。

後者としては、共同研究機関である東北大学乾・岡崎研究室により研究が進められている言論マップ技術のリスクコミュニケーションへの応用が考えられる。具体的には、意見形成の背景にある言論の相互関係を明らかとし、ネット時代のリスクコミュニケーションにおける最適な広報対象と効果的な広報戦略の検討などが考えられる。

ただし、いずれの方向性においても研究には予算を要する。とりわけ、仮説のより詳細な検証のためには、調査対象を増やし、客観性の高い類似意見検索や意見間の類似度の定量化手法の開発が求められる。しかしながら、どれだけコストを掛けて仮説の証明を行ったとしても、研究としての精緻化には貢献するものの、肝心のリスクコミュニケーションの改善には至らない。したがって、限られた研究予算の中での方向性としては、後者の示す自然言語処理のリス

コミュニケーション実務への応用が合理的であろうと考えられる。

## F. 研究発表

1. 論文発表  
なし

2. 学会発表  
なし

## リスクコミュニケーションにおけるインターネット上の 広告効果に関する研究

研究分担者 奥村 貴史

（国立保健医療科学院 研究情報支援研究センター 特命上席主任研究官）

### 研究要旨

現在のように放射能に関する正誤のあいまいな情報や伝聞による健康情報がネットに溢れた場合、信憑性の高い情報を求めるネット利用者は、検索エンジンを利用しより確からしい情報を求めるであろうと考えられる。そこで本研究分担では、ネット時代におけるリスクに関する情報伝達を改善していくために、放射能汚染に関する公的な情報提供サイトに対する検索エンジン上の効率化手法を検討した。まず、公的機関の行う放射能情報提供サイトをリストアップした上で、「検索エンジン最適化」の観点から既存の情報提供サイトの課題を整理した。その上で、利用が拡大している「検索連動型広告」の効果を確認するため、主要な検索エンジンに検索連動型広告を出稿し、その広告効果を測定し効果的な広告手法についての検討を行った。また、以上の検討により、現在の情報提供サイトが利用者のニーズに合致していない可能性が示唆された。今後、検索連動型広告の活用と共に、受け手の特性やニーズに応じたサイト構成やコンテンツの提供を検討する必要がある。

### A. 研究目的

2011年3月に生じた東京電力福島第一原発事故により、環境中に大量の放射性物質が放出された。この事故による公衆の被曝線量は、事故初期の放射線への曝露に関しては未だ解明されていない部分はあるものの、概ね国際放射線防護委員会が勧告する緊急時や現存被曝状況での参考レベルを超えないものと推定されている。しかしながら、前例の無い規模の放射能災害の発生に加え、国や東京電力を初めとする公的機関による情報提供上の問題により、発災直後から現在に至るまで、放射能に関する正誤のあいまいな情報や伝聞による健康情報がインターネットを中心として蔓延する事態が生じている。

リスク情報の適切な提供を目指す「リスクコミュニケーション」は、こうした事態

に対応するため、今までにも、環境汚染による風評被害などに対して実践されてきた。しかしながら、従来のリスクコミュニケーションは、専門家から一般人への一方通行な情報伝達、ないし、専門家と一般人の双方向な交流のいずれかを前提としており、現在のように放射能に関する正誤のあいまいな情報がネットに溢れる状況を想定していなかった。

こうした事態を受け、政府機関や関連学会等も放射能汚染への対策としてインターネット上のさまざまな情報提供を進めている。しかしながら、公的組織には、アクセス数が死活問題となる民間企業のようにアクセス数向上のノウハウが蓄積されておらず、折角の提供情報が十分に活用されていない懸念がある。

そこで、本研究では、ネットが普及した現在におけるリスクに関する情報伝達を

改善していくために、放射能汚染に対する公的組織の情報提供サイトの効率化手法を検討する。その際、信憑性の高い情報を求めるネット利用者の行動として、検索エンジンの利用が想定されるために、まず、検索エンジン経由でのアクセス向上において重要となる「検索エンジン最適化」の観点から、既存の情報提供サイトの課題を整理した。そのうえで、昨今利用が拡大している検索結果に連動した形で広告を配信する「検索連動型広告(リスティング広告)」による情報提供の効率化について検証を行った。

## B. 研究方法

本研究では、まず、公的機関が提供する放射能情報提供サイトを広くサーベイし、官公庁関係 10 サイト、研究所・大学・公益法人 25 サイト、地方自治体 70 サイト、教育・普及啓発 4 サイトの 109 件を抽出し、全体的な傾向と情報提供における課題をまとめた。そのうえで、代表的な 10 サイトを抽出し、サイト構成やタイトル、Meta タグ等、検索エンジン結果に影響を与える点についての問題点を整理し、改善方法の検討を行った。

さらに、インターネットにおける主要な検索エンジンに対して検索連動型広告を出稿し、クリック数などの広告効果を測定すると共に、サイト内での挙動を計測することで、効果的な広告手法についての検討を行った。

具体的な広告としては、国立保健医療科学院生活環境研究部が開設している放射線情報提供サイトを対象として、Yahoo! と Google に対して、それぞれ 4 種の広告を配信した。ケース 1 では、アクセス解析から設定された検索頻度の高いキーワードを設定した。ケース 2 では、「放射能」と「放射線」をメインキーワードとして設定した。ケース 3 では、ケース 2 より得られる関連キーワードをキーワードとして

設定した。最後に、ケース 4 として、ケース 3 のキーワードに加えて広告文章の調整を加えた。これら 4 種類の広告に対して、それぞれの検索エンジン毎に 5 万円、合計 40 万円の広告を配信し、ユーザーの反応を計測した。

以上により、自然検索、並びに、検索連動型広告における課題を明らかにするとともに、アクセス数向上に向けた方策について検討を行った。

(倫理面への配慮)

本研究について、「リスクコミュニケーションに関するインターネット上の広告効果に関する研究」として、国立保健医療科学院倫理審査委員会に研究倫理審査申請を行い、NIPH-IBRA#12038 として承認された。

## C. 研究結果

公的機関による放射線情報提供サイトに関しては、情報提供はしているもののユーザーに適切な導線を提供していない、他サイトとの適切な連携がなされていない等の事例が見られた。そのなかでも、放射線医学総合研究所、日本の環境放射能と放射線、放射線影響研究所の 3 サイトは、ユーザーへの配慮が比較的なされており、サイトとしての完成度が高かった。

次に、これらの 3 サイトに加えて、国立保健医療科学院生活環境研究部が開設している放射線情報提供サイトである「医療での放射線安全の疑問にお答えします」、並びに、東京都健康安全研究センター「放射能 Q&A」、国立環境研究所、農業環境技術研究所、環境省、原子力委員会、日本保健物理学会「専門家が答える Q&A」を加えた 10 サイトを対象に、検索エンジン最適化に向けた解析を試みた。

		表示回数	クリック数	クリック率	平均クリック単価
ケース1	Google	27,369 回	78 回	0.28	283 円
	Yahoo!	75,792 回	95 回	0.13	284 円
ケース2	Google	30,974 回	160 回	0.52	310 円
	Yahoo!	65,739 回	188 回	0.29	257 円
ケース3	Google	103,200 回	537 回	0.52	93 円
	Yahoo!	225,945 回	829 回	0.37	62 円
ケース4	Google	92,535 回	757 回	0.82	67 円
	Yahoo!	229,307 回	1,262 回	0.55	38 円

表 1. 広告手法別訪問者データ

	訪問数	平均PV	平均滞在時間	新規訪問	直帰率
検索連動型広告	3,759	1.71	0:00:58	85.7%	71.2%
自然検索	578	2.98	0:02:15	55.2%	55.0%
その他	124	2.20	0:02:15	67.7%	64.5%

表 2. 経路別訪問者データ

その結果、サイト内 meta 情報についての不備、画像の Alt 情報の不備、キービジュアルの不備、ディレクトリ登録と外部からの被リンク獲得上の課題、マルチデバイスへの未対応問題等が明らかとなった。また、ページ内キーワード比率、アクセス解析を活用したサイト改善、ソーシャルネットワークとの連携による双方向性の確立等についての課題が明らかとなった。それぞれの問題点と説明については、調査報告書に詳述されている。

次に、国立保健医療科学院生活環境研究部が開設している放射線情報提供サイトに対し、ケース 1~4 の広告を行った。その結果、それぞれ表 1 に示すアクセスを得た。また、実験期間中の、広告経由、自然検索経由、並びに、他サイト経由のアクセス統計をまとめたものを表 2 に示す。なお、調査期間の開始直前の 8 日間分をサンプリングしたところ、1 日あたり訪問数 24.5 件、平均ページビュー 3.1、平均滞在時間 2 分 34 秒、新規訪問 57.7%、直帰率 51.0%であった。

表 1 に示されているように、現在のサイトアクセスにおける検索キーワードを用いた広告(ケース 1)から、キーワードと広告文を改善していくのに従って、クリック数、クリック率の増加に加え、クリック単価の改善が生じていることが分かる。また、表 2 より、今回配信した広告により誘導されたユーザーが、自然検索のユーザーよりも平均滞在時間が半分以下であり(58 秒対 2 分 15 秒)、直帰率も高い(71.2%対 55.0%)ことが示された。このデータからは、広告により誘導されたユーザーの関心がサイトの内容と合致していない、ないし、そもそもサイトの内容の魅力が乏しい等のサイト側の問題が示唆される。

#### D. 考察

検索連動広告の配信並びに既存の放射線情報提供サイトの静的な解析により、検索連動広告の利用による露出の改善、並びに、今後の配信における広告キーワードの工夫の必要性が示された。また、利用者がサイトに辿り着くもう一つの導線である自然検

索の表示順位向上に向けた工夫についても、検索エンジン最適化の観点からの既存サイトの課題が明らかとなった。一方で、これらの改善を行ったとしても、サイトそのものが利用者のニーズに適合していない限り、利用者は誘導されてもコンテンツは閲覧せず、そのままサイトを離脱することが明らかとなった。したがって、今後の課題として、コンテンツそのものの改善を通じた直帰率の低減を目指す必要がある。

コンテンツの改善に際しては、各サイトは、具体的な利用者を想定した上でのサイト構成やコンテンツの準備が求められる。たとえば、情報提供の対象が若年男性と主婦とでは、求めている情報の分野も提供する情報の詳しさも異なってくるだろう。利用者によっては客観的な数値データや各種のグラフを望むこともあれば、たとえ話による説明や単純化した図を望むことも考えられる。こうした工夫は、サイトの露出や売上げが厳しく追い求められる EC サイト(通販サイト)では常識的なものであるにも関わらず、公的機関による情報提供においては省みられることが少ない。

とりわけ、公的機関の情報は、所掌毎にカテゴリに分けられ、高度に構造化されていることが多い。これは情報を作成する側の都合でもあるが、情報は構造化されている方が情報を主体的に探す利用者にとっても便利でもあるため、問題とされることは少ない。一方で、受動的な利用者は、テレビのように作り手側がストーリーを提供することを期待していることが少なくない。実際、通販サイトにおいては、売り手が一方的にストーリーを伝える縦長レイアウトによる商品紹介の方が商品の売上げも多くなることが経験的に知られている。そのために、主婦が利用するウェブサイトとしても、縦長ページが主体のさまざまな業者サイトが無秩序に並ぶ楽天市場が、商品情報が高度に構造化されている Amazon.com

の倍近い人気を得ている等の統計がある。

したがって、放射線の情報提供においても、客観的な情報を構造化して提供するだけでなく、ストーリーを有した縦長ページなど、ユーザー毎の特性に即した情報提供を試行することにより、利用者ニーズにより応えることができる可能性がある。また、広告のキーワードや広告文章毎に誘導先として構造化されたページと非構造化されたページを切り替えるなどの工夫により、より効果的な情報提供が可能となる可能性もある。こうした試みの効果については、構造化された情報提供と縦長ページによる情報提供との比較研究により容易に検証も可能であるため、検討の余地がある。

## E. 結論

ネットの爆発的な普及により、多くの国民にとってネットの存在は無視し得ないものとなった。本研究班の他研究分担の調査においても、公的機関による放射線情報の提供には期待が高いことが明らかとなっており、ネットを活用した効率的な情報提供が望まれている。そこで、本研究分担では、検索エンジンの有効活用を通じた、放射能汚染に対する公的組織の情報提供サイトの改善に向けた検討を行った。

検索エンジンからのアクセス向上において必要となる検索結果の改善については、「検索エンジン最適化」の観点から、既存の情報提供サイトについて多くの課題を明らかとすることが出来た。また、昨今利用が拡大している「検索連動型広告」の広告効果研究では、キーワードと広告文の改良によりより効率的な広告配信が行えることが明らかとなった。今後、受け手の特性やニーズに応じたサイト構成やコンテンツの工夫、広告のカスタマイズとその効果の検証が望ましい。

本研究は検索エンジンを通じた情報提供



の改善について検討したが、昨今利用が拡大している Twitter や Facebook 等のソーシャルネットワークを利用した公的機関による広報活動については考慮をしていない。これらの手段を用いた広報により、より利用者に近い位置でのより細かな情報提供が可能である。さらに、ネットでの情報提供サイトをテレビやラジオにより広報するような情報提供は、露出の高さとターゲットを絞った情報提供を両立させられる可能性がある。来年度以降、これらのテーマについても検討を進めることにより、公的機関による情報提供をさらに改善していける可能性がある。

## F. 研究発表

### 1. 論文発表

なし

### 2. 学会発表

なし

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
鍋島啓太, 渡邊研斗, 水野淳太, 岡崎直観, 乾健太郎.	訂正パターンに基づく誤情報の収集と拡散状況の分析.	自然言語処理	20(3)	印刷中	2013
山口一郎.	食品中の放射性物質の新たな基準値を考えるための基礎知識.	ファルマシア	49(1)	32-36	2013
山口一郎.	医療で用いられる放射線, 放射性物質の基礎知識.	安全医学	8(1)	10-20	2012

# 訂正パターンに基づく誤情報の抽出と集約

鍋島啓太<sup>†</sup> 水野淳太<sup>†</sup> 岡崎直観<sup>†‡</sup> 乾健太郎<sup>†</sup>  
東北大学大学院 情報科学研究科<sup>†</sup> 科学技術振興機構 さきがけ<sup>‡</sup>  
{nabeshima, junta-m, okazaki, inui}@ecei.tohoku.ac.jp

## 1 はじめに

2011年3月に発生した東日本大震災では、ソーシャルメディアは有益な情報源として活躍した。その一方で、「イソジンを飲むと放射線予防になる」に代表されるような誤情報の拡散が問題となった。このような誤情報の中には人々の安全に関わるものも多く、誤情報に対する注意喚起を低コストで実現する仕組みが必要である。

誤情報の検出に関する研究は多くされている [1, 2, 3, 4]。Qazvinianら [1] は、誤情報に関連するツイート群から、誤情報に言及しているツイートと、誤情報に言及していないツイートに分類し、さらに誤情報に言及しているツイート群を、誤情報を支持するツイートと否定するツイートに分類する手法を提案した。梅島ら [2] は、訂正を明示する表現（「デマ」など）を含むツイートを収集し、各ツイートが特定の情報を訂正しているか、訂正していないかを識別する二値分類器を構築した。これらの先行研究は、ツイートの本文を単位とし、誤情報を含むか、もしくは特定の情報を訂正しているかどうかを認識することに注力しており、ツイート本文中から誤情報の箇所をピンポイントで特定しているわけではない。

本論文では「〇〇というのはデマ」などの誤情報を訂正する表現（以下、訂正パターン）に着目し、東日本大震災後1週間の全ツイートから誤情報を自動的に収集する手法を提案する。提案手法により、既存のまとめサイトに収録されている60件の誤情報の約半数を再現でき、さらにまとめサイトに収録されていない22件の誤情報を獲得することができた。

## 2 提案手法

図1に提案手法の流れを示す。手順は大きく4つに分けられる。以降では、各ステップについて説明を行う。

**ステップ1 被訂正フレーズの抽出:** ステップ1では、ツイート本文から被訂正フレーズを見つけ出す。被訂正フレーズとは、「イソジンは被曝を防げるというのはデマだ」の下線部のように、「デマ」や「間違い」といった訂正表現で打ち消されている箇所のことである。被訂正フレーズと訂正表現は、「という」や「のような」といった連体助詞型機能表現で繋がれており、被訂正フレーズに続く表現を「訂正パターン」と呼ぶ。人手で作成した368個の訂正パターンのいずれかにマッチするツイート本文に対して、文頭から訂正パターンの直前までを被訂正フレーズとして抽出する。本ステップをツイート全体に適用し、抽出した被訂正フレーズの集合を  $D$  とする。

**ステップ2 キーワードの抽出:** 前節で抽出された被訂正フレーズには、「昨日のあれはデマだ」の「昨日のあれ」のように、具体的な情報に言及していないフレーズも含まれている。これらは誤情報としては不十分であるため、

取り除く必要がある。そこで、被訂正フレーズ中の単語が訂正パターンとよく共起しているかどうかを調べる。具体的には、ある語  $w$  がツイートで言及されるとき、その語が被訂正フレーズ集合  $D$  に含まれる条件付き確率、

$$P(w \in D|w) = \frac{w \text{ が訂正パターンと共起するツイート数}}{w \text{ を含むツイート数}} \quad (1)$$

を算出し、確率が高い上位500単語を誤情報のキーワードとして選択する。

**ステップ3 キーワードのクラスタリング:** 被訂正フレーズには、「コスモ石油の火災により有害物質を含む雨が降る」と「コスモ石油の爆発は有害だ」のように、同一の情報に言及しているが、表現や情報量の異なるフレーズが含まれている。誤情報を重複なく抽出するために、これらをまとめる必要がある。そこで、ステップ2で抽出されたキーワードをクラスタリングする。キーワード間の距離（類似度）として、キーワードと文内で共起する内容語（名詞、動詞、形容詞）を特徴量とした文脈ベクトルのコサイン距離を用いた。文脈ベクトルの特徴量には、キーワードと各単語との共起度合いを測定する尺度である自己相互情報量を用いた。クラスタリング手法として最短距離法を用いた。各クラスタにおいて、ステップ2の条件付き確率が高いものを代表キーワードとする。

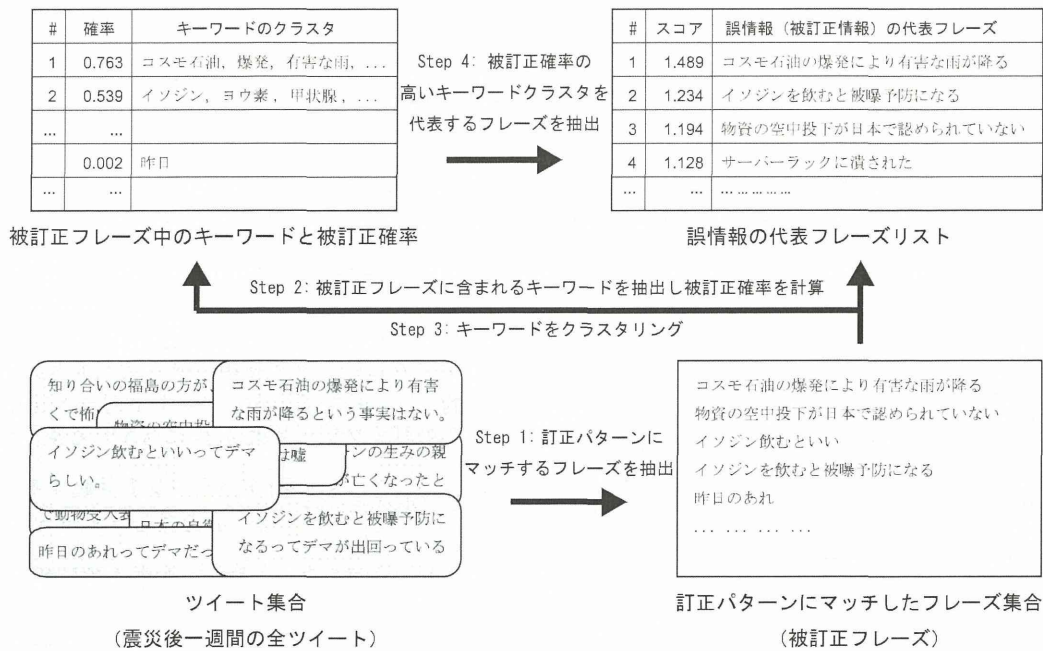
**ステップ4 代表フレーズの選択:** 前ステップで得られた各クラスタに対し、そのクラスタ中のキーワードを含む被訂正フレーズの中で代表的なものを選択し、誤情報として出力する。誤情報を過不足なく説明できる被訂正フレーズを選択するため、以下の式でスコアを計算する。

$$\text{score}(s, t) = \text{hist}(\text{len}_s, t) \times \sum_{w \in C_s} \text{PMI}(t, w) \quad (2)$$

ここで、 $s$  は被訂正フレーズ、 $t$  は誤情報クラスタを代表するキーワード、 $C_s$  は  $s$  中の内容語の集合、 $\text{len}_s$  は被訂正フレーズ  $s$  の単語数を示す。  $\text{hist}(l, t)$  は、最重要キーワード  $t$  を含み、かつ単語数が  $l$  である文の出現頻度、  $\text{PMI}(t, w)$  は  $t$  と単語  $w$  の自己相互情報量を示す。式 (2) は、キーワードとよく共起する内容語を多く含み、かつ標準的な長さの被訂正フレーズに対して、スコアが高くなるように設計されている。すなわち、  $\text{hist}(\text{len}_s, t)$  は、最重要キーワードを含むフレーズの中で典型的な長さのフレーズに高いスコアを与え、極端に短いフレーズ・長いフレーズに対して低いスコアを与える補正式である。

## 3 実験

評価実験では、東日本大震災時のツイートデータを用いて誤情報の抽出を行い、その精度と再現率を測定した。



### 3.1 実験設定

誤情報の抽出元となるコーパスには、東日本大震災ビックデータワークショップ<sup>1</sup>でTwitter Japanから提供された2011年3月11日9時から2011年3月18日9時までの179,286,297ツイートを利用した。評価実験の正解データとして、誤情報を人手でまとめた以下の4つのウェブサイト<sup>2</sup>に掲載されている事例のうち、Twitterデータの投稿期間内に発信されたと判断できる60件の誤情報を用いた。提案手法で抽出された誤情報の正否は、同等の内容が60件の正解データに含まれるかどうかを一件ずつ人手でチェックを行うことで判定した。また、これらの4つのまとめサイトに収録されていないが、誤情報であると判断できるものもある。そこで提案手法が抽出した情報が正解データに含まれなかった場合は、人手で調査を行い、実際には誤情報だったのか判断した。本研究の目的は、誤情報を網羅的に抽出することであるので、抽出した誤情報のうち、同じ内容と判断できるものが複数ある場合、正解は1つとした。評価方法について、提案手法はスコアの高い順にN件まで出力可能であるため、Nを変化させたときの精度、再現率を計測した。

### 3.2 実験結果

評価結果を表1に示す。Nが100のとき、提案手法が抽出した情報のうち、正解データにも存在する情報は3割である。さらに、今回の正解データには含まれないが、誤情報と判断できる事例が約2割あり、提案手法は約5割の適合率で誤情報を抽出できた。不正解だった事例のうち、約半数は同じ誤情報を別のフレーズで表現したものが占めるため、提案手法が抽出する誤情報の約7割は正解と見なすことができる。

<sup>1</sup><https://sites.google.com/site/prj311/>

<sup>2</sup>収集したサイトは以下の通り

<http://www.kotono8.com/2011/04/08dema.html>

<http://d.hatena.ne.jp/seijotcp/20110312/p1>

<http://hara19.jp/archives/4905>

<http://matome.naver.jp/odai/2130024145949727601>

表 1: 抽出された誤情報の精度・再現率

N	精度 (4 サイト)	精度 (人手判断)	再現率
25	0.44(11/25)	0.64(16/25)	0.18(11/60)
50	0.34(17/50)	0.58(29/50)	0.28(17/60)
75	0.33(25/75)	0.56(42/75)	0.42(25/60)
100	0.30(30/100)	0.52(52/100)	0.50(30/60)

## 4 おわりに

本研究では、誤情報を訂正する表現に着目し、誤情報を自動的に収集する手法を提案した。実験では、誤情報を人手でまとめたウェブサイトから取り出した誤情報のリストを正解データと見なして評価した。抽出された情報の中には、まとめサイトに掲載されていない誤情報も存在し、提案手法は誤情報の自動収集に有用であることが分かった。今後は、訂正パターンの拡充や被訂正フレーズのスコアリングの改良を進め、誤情報抽出の性能を向上させるとともに、リアルタイムでの誤情報獲得に取り組む予定である。

### 謝辞

本研究は、文部科学省科研費(23240018)、文部科学省科研費(23700159)、およびJST戦略的創造研究推進事業さきがけの一環として行われた。データを提供して頂いたTwitter Japan 株式会社に感謝いたします。

### 参考文献

- [1] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In *Proc. of EMNLP 2011*, pp. 1589–1599, 2011.
- [2] 宮部真衣, 梅島彩奈, 灘本明代, 荒牧英治. 流言情報クラウド: 人間の発信した訂正情報の抽出による流言収集. 言語処理学会第18回年次大会, 2012.
- [3] 藤川智英, 鍛冶伸裕, 吉永直樹, 喜連川優. マイクロブログ上の流言に対するユーザの態度の分類. 言語処理学会第18回年次大会, 2012.
- [4] 鳥海不二夫, 篠田孝祐, 兼山元太. ソーシャルメディアを用いたデマ判定システムの判定精度評価. デジタルプラクティス, Vol. 3, No. 3, pp. 201–208, 2012.