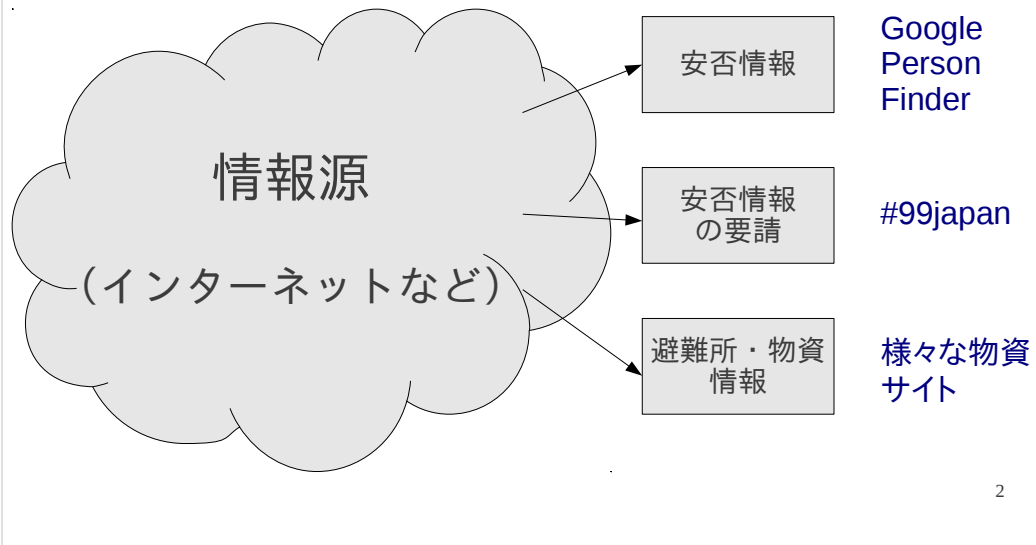


震災後のテキストデータからの
情報抽出ツール開発
～ webigator チーム～

ありがとうございます。
今回の発表は震災後のテキストデータから情報の抽出を
助けるツール webigator について話します。

東日本大震災後の情報抽出・集約



このプロジェクトをしようと思った理由は、東日本大震災の後にインターネットで飛び交っている情報を集約している活用を見てきたからです。

例えば、安否情報を Google Person Finder や 99japan など、避難所や物資の情報を様々なサイトに集約する活用がありました。

プロジェクトの概要

- 震災後、情報を集約するニーズが高い
- 手伝いたい人がたくさんいる
- 関係のない情報が氾濫

これを見て分かったことは震災後、情報を集約するニーズがとても高いことや、このような活動に参加したい人が非常に多いことです。

ただ、ネット上にこのような有用な情報はたくさんありますが、関係のない情報もたくさんあります。

この関係のない情報を全部一通り見ることが結構時間がかかり、この情報を事前に弾くことができれば、より効率的に安否や物資などの情報をまとめられのではないかと考えました。

プロジェクトの概要

- 震災後、情報を集約するニーズが高い
- 手伝いたい人がたくさんいる
- 関係のない情報が氾濫

関係のない情報を見る手間を省き、
集約したい情報を効率良く提示する
システムの開発に取り組んだ

4

そこで、今回のプロジェクトは我々の専門である言語処理と機械学習の技術を使って、関係のない情報を見る手間を省いて、集約したい情報を効率良く提示するシステムを目標にしました。

また、複数のユーザが同時に使えるようにして、初めて使った人でもそこそこ分かりやすく使えるページにしてみました。

ツール利用の流れ

次はシステム全体の利用の流れをこのスライドに示しています。

ツール利用の流れ

1) 検索したいキーワードを選択

“ 給水 ” “ 避難所 ” “ 炊き出し ”

6

まずは、集約したい情報に関するキーワードを何個か入れます。

例えば、物資情報を調べたいなら、「給水」「避難所」「炊き出し」などのキーワードを入れます。

ツール利用の流れ

1) 検索したいキーワードを選択

“ 給水 ” “ 避難所 ” “ 炊き出し ”

2) システムがテキストを提示

○○ 県の△△市の避難所は□□と◇◇です!

7

その次、システムがこのキーワードに基づいてテキストを検索して提示します。

今のところは Twitter からのデータを対象にしていますが、少し変更することでほかのテキストも取り組めます。

ツール利用の流れ

1) 検索したいキーワードを選択

“ 給水 ” “ 避難所 ” “ 炊き出し ”

2) システムがテキストを提示

○○ 県の△△市の避難所は□□と◇◇です!

3) 提示されたテキストの情報を集約

この提示されたテキストの中に有用な情報、例えばこの場合だと支援物資に関する情報、があれば、これを一覧やウェブサイトに入れておきます。

ツール利用の流れ

- 1) 検索したいキーワードを選択

“ 給水 ” “ 避難所 ” “ 炊き出し ”

- 2) システムがテキストを提示

○○ 県の△△市の避難所は□□と◇◇です!

- 3) 提示されたテキストの情報を集約

- 4) 探している情報が含まれたか? + / -

○○ 県の△△市の避難所は□□と◇◇です!

9

入れ終わった時に、このテキストが役に立ったなら「+」を、役に立たなかったなら「-」を押します。

ツール利用の流れ

- 1) 検索したいキーワードを選択

“ 給水 ” “ 避難所 ” “ 炊き出し ”

- 2) システムがテキストを提示

○○ 県の△△市の避難所は□□と◇◇です!

- 3) 提示されたテキストの情報を集約

- 4) 探している情報が含まれたか? + / -

○○ 県の△△市の避難所は□□と◇◇です!

- 5) システムが学習し、より正確な情報を提示¹⁰

そして、この+と-の情報を使って、システムが役に立ったテキスト、役に立たなかったテキストのパターンを学習します。

このパターンに基づいて、次のテキストを提示した時に、システムがより役に立ちそうな結果を提示してくれるようになり、役に立たないテキストを読む時間が少なくなります。

タスク一覧画面

webigator デモ

説明

+ タスクの作成

以下は現在進行中のタスクです。タスクに参加したい場合は「参加」ボタンをクリックしてください。
鍵マークのついたタスクへ参加するにはパスワードが必要です。

タスクID	キーワード		
12382101	給水 炊き出し 充電	 参加	 設定
491991293	安否 確認 できました	参加	 設定

11

ツールの実際の画面はこのようになっています。

上に新しいタスクを作成するボタン、下の方に現在進行中のタスクが載っている一覧があります。

ここでタスクは「1種類の情報」という意味ですが、例えばこの画面だと現在進行中のタスクを見ると「給水、炊き出し、避難所」というキーワードを持った物資情報に関するタスクと、「探しています、安否」などの安否情報に関するタスクがあります。

タスク作成画面

webigator デモ

説明

必要な項目を入力し、「登録」ボタンをクリックしてください。

検索キーワード (必須)	<input type="text"/> 複数のキーワードを指定する場合は、空白で区切ってください。
管理用パスワード (必須)	<input type="password"/> 登録したタスクの設定を変更する場合に使用します。
作業用パスワード	<input type="password"/> パスワードを設定した場合、タスク参加時にパスワードの入力を求められます。 パスワードを設定しない場合は、誰でもタスクに参加できます。

登録

削除

「タスク作成」のボタンを押すと、現在進行中のタスク以外のタスクを作成することができます。
例えば、上のキーワードに「道路、鉄道、徒歩」のようなキーワードを入れて交通情報に関するタスクを作れます。
また、パスワードを入れて作業する権利をパスワードが分かっている人に制限することができます。

webigator デモ

作業画面

説明

現在のキーワード 給水 安否

有用なものは“+”のラベルを、有用でないものは“-”のラベルを選択してください。
判断のつかないものは、そのままかまいません。

ラベル	テキスト	ツイートID
+ -	RT @muhmue: 【被災者の皆さんへ】盲点でした。NHK教育のみ、被災者の方への有益な情報を流しています。給水所・無料電話設置所・営業中店舗・安否情報などが流され、計画停電情報も北関東中心に報道されます。そして合間には子供向け番組も放送されます。悲惨なシーンも一切 ...	47459235637895168
+ -	自然災害はどうもすることが出来ない・・・自然に治まるのを待つだけか・・・それには多数の安否者が出る。ガキツカ見てる場合ちゃうで俺	46690348545753088
+ -	RT @kahoku_shimpo: 生活情報 / ごみ片付け・給水手伝い窓口 仙台ボランティアセンター開設 http://bit.ly/e5crSn	47947288718483456
+ -	RT @ienokettosyo: 泉区で本日給水がうけられるのは17か所 将監中央小学校 南光台市民センター 高森小学校 桂小学校 七北田小学校 虹の丘小学校 長命ヶ丘小学校 将監小学校 向陽台中学校 泉区役所 住吉台小学校 館中学校 黒松小学校 寺岡中学校 加茂小学校 ...	47590447413071873
+ -	@volvox314 給水があるからなんとかなってるけどね^^ ありがとう、早く元の生活をしたいものだ	47431038938517505

13

ラベルを投稿

最後に、実際の作業画面です。

上の方にタスクのキーワードが書いてあります。

下に、提示されたテキストの一覧があり、左側に+と-のボタンで役に立ったかどうかをシステムに教えます。

最後に「ラベルを投稿」を押し、システムがこのラベルに基づいてパターンを学習し、新たにテキストを提示してくれます。

ツールの入手

- オープンソースで公開：
 - <http://www.phontron.com/webigator>
 - デモも公開中
- 協力者募集中！
 - サーバ・Web アプリの機能拡張
 - ユーザからのコメント也大歓迎
 - 興味のある方は neubig@gmail.com へご連絡ください
- 開発者：
 - Graham Neubig - 計算サーバ/Web アプリの開発
 - 村上慎吾 - UI/Web アプリの開発

14

このツールを誰でも使えるようにこのページからオープンソースで配布しています。

デモも公開中ですので、ちょっと試すこともできます。

ただ、これからの課題もたくさんありますので、サーバとWeb アプリの昨日を拡張してくれる方、ツールを使ってコメントをしていただくユーザがいらっしゃるなら、ぜひご連絡ください