

英文Web文書中の ユーザが知らない単語を 予測する語義注釈システム

旧題：語義注釈システムの単語クリックログからの言語能力情報の抽出

2009/09/30

東京大学

○江原 遥, 二宮崇, 中川 裕志

<http://www.socialdict.com/>

既存の語義注釈システム

pop辞書 (2001)

pop-upで表示する

Han Chinese groups paramilitary

riots

Los Angeles Times - Davi

A group of Han Chinese ca *不完全一致

China. State police and paramilitaries deploy by the thousands in a bid to

(団体など)軍補助的な、準軍事的な

popIn (程, 2008)

なぞって選択した文字列に対して訳を埋め込む

政府は2005年から、温暖化防止対策(Global Warming Prevention)の一環として夏のビジネスで軽装を勧めるクールビズ運動を開始。官民を問わず定着しつつあった流れが、今回の選挙結果を受けて変わるのだろうか。

提案

予測機能を追加する:

知らない語をいちいちクリックするのは面倒。Web文書中のユーザが知らなさそうな語を予測し、予め注釈をつけれると嬉しい。

黄色: ユーザが知っている(既知)と判定された語

赤色: ユーザが知らない(未知)と判定された語

Military **chief** Gen Victor Ibrado **flew** **by** **helicopter** **over** **suburbs**(**郊外**, **近郊**) of Manila **on** Sunday **to** **see** **for** **himself** **drenched** **survivors** **marooned** **on** **top** **of** **half-submerged** **buses** **and** **on** **rooftops**.

TV **footage** **showed** **some** **survivors** **clinging**(**くっついて離れない**, **固守する**, **くっつく**, **執着する**) **to** **high-voltage** **power** **lines**.

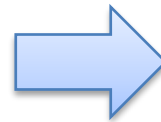
提案

Military chief Gen Victor Ibrado flew by helicopter over suburbs(郊外, 近郊) of Manila on Sunday to see for himself drenched survivors marooned on top of half-submerged buses and on rooftops.

TV footage showed some survivors clinging(くっついて離れない, 固守する, くっつく, 執着する) to high-voltage power lines.

rooftops. (1.屋根)

単語クリックログ



(0[未知], u_1 , "rooftop")

黄色: ユーザが知っている(既知)と判定された語

赤色: ユーザが知らない(未知)と判定された語

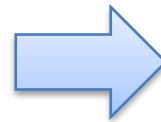
提案

Military chief Gen Victor Ibrado flew by helicopter over suburbs(郊外, 近郊) of Manila on Sunday to see for himself drenched survivors marooned on top of half-submerged buses and on rooftops.

TV footage showed some survivors clinging(くっついて離れない, 固守する, くっつく, 執着する) to high-voltage power lines.

clinging to high-voltage

単語クリックログ



(1[既知], u_1 , "cling")

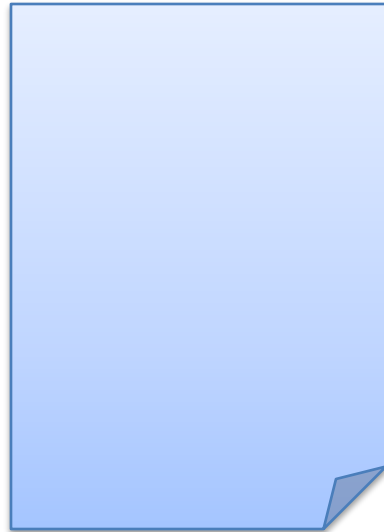
黄色: ユーザが知っている(既知)と判定された語

赤色: ユーザが知らない(未知)と判定された語

利点

- クリックが不可能/難しい場合

印刷時(不可能)



PDAなど(難しい)



まとめ

- 英文Web文書中のユーザが知らない単語を予測して、自動的に語義の注釈を行うシステムを作成した
 - コーパス別の単語帳の自動作成
 - ACLやEMNLPの論文の頻出単語のうち自分が知らない語は？
 - TOEFLなどの言語テストで使用されている項目反応理論をロジスティック回帰に帰着させて使用
 - 有名なクラウド環境Google App Engineを使用
 - **デモをやっています！詳しくはポスターで！**
- <http://www.socialdict.com/>