



# 大規模言語資源, フラゲ!

橋本力<sup>1</sup> 鳥澤健太郎<sup>1</sup> 黒田航<sup>2</sup> Stijn De Saeger<sup>1</sup>  
村田真樹<sup>3</sup> 風間淳一<sup>1</sup>

<sup>1</sup> 情報通信研究機構(NICT) <sup>2</sup> 杏林大学 <sup>3</sup> 鳥取大学

[受賞論文]

WWW からの大規模動詞含意知識の獲得

橋本力, 鳥澤健太郎(NICT), 黒田航(杏林大学), Stijn De Saeger(NICT), 村田真樹(鳥取大学), 風間淳一(NICT)  
情報処理学会論文誌, Vol.52, No.1, pp.293-307 (2011)

このたび標記の論文で本会論文賞をいただくことになりました。ご協力いただいたすべての方々に感謝申し上げます。

本論文は、計算機による推論や質問応答等に欠かれない、『購入する』が『入手する』を含意する等の動詞含意知識を膨大な Web 文書から自動獲得する新手法を述べたものです。これにより、たとえば「新作ゲーム A を入手できるのはどこ?」という質問に対し、「ショップ B で A をフラゲ!」を Web から取得して回答できるようになります。実用に耐える技術としては「フラゲする」のような新語も大量に扱わなくてはなりません。我々は 1 億の Web ページから抽出した 20 億以上の動詞の対を対象に知識獲得しましたが、この規模になると大規模計算機クラスターが必須です。でも心配いりません。我々の獲得結果を人手チェックしたもの(約 12 万対、商用利用可)を ALAGIN フォーラムから入手できます。

最近「帳消しにする」と「チャラにする」等のフレーズの同義関係、「ガンを破壊する」と「ガンを進行させる」等の矛盾関係、「インフレが発生する」と「円安になる」等の因果関係も Web から 100 万個規模で獲得し、さらに Web にも明示的に書かれていない因果関係の仮説もやはり 100 万個規模で自動的に生成しています。我々にご一報いただければフラゲできるかも?

(2012 年 4 月 15 日受付)



橋本力 (正会員) ch@nict.go.jp

京都大学情報学研究科研究員, 山形大学理工学研究科助教を経て、情報通信研究機構研究員。博士(言語科学, 情報学)。言語処理学会論文賞, 言語処理学会年次大会優秀発表賞。

鳥澤健太郎 (正会員) torisawa@nict.go.jp

1995 年東京大学大学院博士課程中退。同年同大学院助手。北陸先端科学技術大学院大学助教授を経て、現在、情報通信研究機構情報分析研究室室長。博士(理学)。日本学術振興会賞等受賞。

黒田航 kow.kuroda@gmail.com

(独) 情報通信研究機構専攻研究員, 京都工芸繊維大学, 早稲田大学総合研究機構を経て、現在、杏林大学教員。京都大学から人間・環境学博士を取得。言語学と自然言語処理を融合する研究に従事。

Stijn De Saeger stijn@nict.go.jp

2006 年北陸先端科学技術大学院博士課程修了。博士(知識科学)。情報通信研究機構専攻研究員を経て、現在同機構主任研究員。言語処理学会年次大会優秀発表賞等受賞。

村田真樹 (正会員) murata@ike.tottori-u.ac.jp

1997 年京都大学大学院博士課程修了。博士(工学)。同年、京都大学にてリサーチ・アソシエイト。1998 年通信総合研究所入所。情報通信研究機構主任研究員を経て、2010 年鳥取大学教授。

風間淳一 (正会員) kazama@nict.go.jp

2004 年東京大学大学院情報理工学系研究科博士課程修了。博士(情報理工学)。同年、北陸先端科学技術大学院大学情報科学研究科助教。2008 年より情報通信研究機構。専門は自然言語処理。