

Title	Webでの文章入力時における情報補完
Author(s)	中村, 和正
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3589
Rights	
Description	Supervisor:鳥澤 健太郎, 情報科学研究科, 修士

Web での文章入力時における情報補完

中村 和正 (410092)

北陸先端科学技術大学院大学 情報科学研究科

2007 年 2 月 8 日

キーワード: 文章入力支援, AJAX, 属性, 属性値.

本研究では, ユーザが文書を作成する際に, 作成する文章に関連する情報を Web から自動的に収集し, 提示することにより, 文書の作成支援を行うシステムを提案する.

近年コンピュータを用いて文書を作成する際には, 作成しようとする文章の参考となる関連情報を Web から収集する機会が非常に多い. 例えば Blog で, その日に自分が関与した事物(訪れた場所, 買った商品, 参加したイベントなど)に関する意見や情報の記述を行う際にも, 記述の正確性・詳細性を高めるため 1) 事物の正式名称を確認したり, 2) 事物に関する第三者の意見を参考にしたり, 3) 訪れた場所のアクセスの方法, イベントの正確な日時, 商品の仕様など事物の様々な側面(属性)の情報を確認したり付加したりするなど情報要求が頻繁に生じる. このような情報要求が生じると, ユーザは能動的に商用の全文検索エンジンを利用するなどして, 書く作業を中断して参考情報を収集する必要がある.

本研究では, ユーザが文書を作成する際に頻繁に発生する情報要求に伴う「調べる」作業を極力抑えるために, 1) ユーザの情報要求をこれまで知識獲得に使われていた語彙統語パターン [1] を利用して発見し, 2) その情報要求タイプに応じた知識を Web からオンデマンドで獲得し, ユーザに提示することを目指す. ユーザの情報要求の発見および, Web から得られた知識の提示は, AJAX を用いることでシームレスに行い, 極力ユーザの手間を発生させないように工夫する. ユーザの情報要求は, 実際には様々なタイプのものが存在すると思われるが, 本研究では事物として具体物(例えばドイツやタイタニックなど)である固有名詞に焦点を当て, 吉永と鳥澤による既存研究 [2] を用いて, 具体物の様々な側面, すなわち属性とその属性値を獲得し, ユーザに提示する.

これにより, 例えば「日本酒である太平山の原料米」に関して文章を作成する場合, ユーザがそれを知らなかったとしても本システムが「太平山の原料米は山田錦である」と情報を提示することにより, Web を調べることなくそれを参考にスムーズな文章作成を続行することができるようになる.

本研究で開発するシステムの処理の流れは以下になる。

- 1) ユーザが本システムのクライアント上のエディタ画面から文章を入力
例：「ラーメン屋の真打の住所は」
- 2) クライアントは自動的にバックグラウンドで一定間隔に本システムのサーバーへ入力中の文章を送信
- 3) サーバーが受信したデータからユーザの情報要求を検知できた場合、情報抽出エンジンから関連情報を取得
- 4) 関連情報が取得できた場合、サーバーはクライアントへその関連情報を送信
例：「石川県金沢市大額3丁目184」
- 5) クライアントが関連情報を受信すると、エディタ画面内に表示
- 6) ユーザは表示された関連情報を参考に文章入力を続行

例えば Blog を書いているユーザが「今日は学校の近くのラーメン屋真打に行きました。真打の住所は、」という文章を入力しているとする。ここでユーザが真打の住所を知らなかったとしても、本システムが「真打の住所」を情報要求として検知し、真打の住所に関する情報を Web から発見し、「石川県金沢市大額3丁目184」という情報を画面内に提示することで、ユーザの文章作成を支援する。このように文章入力時において情報要求が発生したとしても、検索にかかる手間をバックグラウンドでサーバーに任せることができるため、従来のように改めて Web ブラウザを起動し、Web 上の検索エンジンにアクセスし、必要な情報を探すといった「調べる」作業のために文章入力を中断させられることがなくなり、文章入力を続行することができるようになる。

本システムは情報抽出エンジンとして、吉永と鳥澤による属性・属性値抽出システム [2] を用いている。本システムはユーザの入力文から語彙統語パターンを用いて、検知された具体物とその属性に関して情報要求を表現するクエリを生成し、属性・属性値抽出システムの入力として与えることで具体物の属性・属性値、つまり情報要求に対して参考となる情報を獲得する。

本稿では、入力中の文章に関連した情報を AJAX と既存の情報抽出エンジンを用いて提示することで、ユーザへの文章作成を支援する手法の1つとして、本システムを提案した。本システムの提示する参考情報が、ユーザにとって本当に参考になるか評価実験を行ったところ、情報要求を検知できた内、50%以上の確率で何らかの情報を提示することができ、その内半数近くの情報が実際に参考となったことが確認できた。また、吉永と鳥澤の論文にあるように今後の精度向上も見込めることから、十分に文書作成の支援ができる情報抽出システムになるものと期待している。

以上から、本研究の目的である“文章作成の中断となるような「調べる」作業を発生させることなく、スムーズに文章を作成することへの支援が行えると考えている。今後は、入力中の文章から検知する語彙統語パターンを増やすことで、ユーザの様々なタイプの情

報要求に対応し，情報要求のタイプに応じて適切に情報抽出エンジンを使い分け，多岐に渡る情報の提示を行いたいと考えている．

参考文献

- [1] Marti A. Hearst, Automatic acquisition of hyponyms from large text corpora. In Proc. of COLING, pp.539 - 545, 1992.
- [2] 吉永直樹, 鳥澤健太郎, Web からの具体物の属性・属性値情報の自動獲得, 言語処理学会第 13 回年次大会発表論文集, 2007.
- [3] 安藤まや, 関根聡, 石崎俊. 定型表現を利用した新聞記事からの下位概念単語の自動抽出. 情報処理学会研究報告, 2003-NL-157, pp.77-82, 2003.