

インタラクティブパラフレーズ

Interactive Paraphrasing of Online Documents

東中 竜一郎 (rh@sfc.keio.ac.jp)

慶應義塾大学

政策メディア研究科

長尾 確(knagao@jp.ibm.com)

日本アイ・ビー・エム株式会社

東京基礎研究所

現在、さまざまなオンラインドキュメントを誰もが簡単にネットワークを介して手に入れられるようになったが、それらのドキュメントはさまざまな専門分野に関連し、語彙も多様なため、ユーザ（閲覧者）にとっては必ずしも分かりやすい単語だけで構成されているとは限らない。そこで、ドキュメントがユーザにとって理解するのが困難な単語を含む場合、どのようにその単語を分かりやすく提示し、ユーザの理解を促進できるかが問題となる。

コンテンツに関するメタデータのように、あるドキュメントに対して、そのドキュメントに関連する情報を付加することはマークアップ言語等を利用することで可能である。そこで、ドキュメントの特定の単語にメタ情報としてその単語の意味情報を追加することで、高度な自動辞書引きを可能にし、ユーザが分からない単語を画面上での簡単な操作で理解可能な単語に言い換える（パラフレーズする）ことを可能にする。さらに、言い換え後の単語でも理解が促進されない場合、ユーザのさらなるインタラクションにより何回でも、インクリメンタルに言い換えができる仕組みを提案する。

1. はじめに

Web 上にさまざまな文書が存在するがそれらは多様な内容と形式を持つために文書によっては背景知識を持たない人にとって分かりにくかったり、意味が取れなかったりすることがある。専門用語や、一部でだけ通用するような単語を含む文書であると他の分野に携わっている人にとっては非常に難解なものになる。これでは Web の発達によって様々な文書にアクセスできる状況になったとはいえ、本当に文書が共有されているとは言えない。誰にとってもある程度理解可能な文書を提供することが可能になれば Web の価値はより高まるのではないかと考える。

難解な単語の理解支援として、オンライン辞書などが利用される。マウス辞書のようにオンライン文書に統合され、単語にマウスカーソルを合わせるとその定義文がポップアップするというものも一般的になった。しかし、従来のマウス辞書では語義が解決できておらず、複数語義のある単語について辞書引きを行った場合、その単語に関するすべての語義を表示してしまう。これは使用される辞書が非常に一般的なためと、文書のコンテキスト

からその単語がどの意味に使われているのかを決定することができないからである。従って、語義の候補が複数個ある場合、閲覧者はその中から適切な候補を選ばなくてはならない。

アノテーションとよばれる付加データをオンライン文書中の要素に関連付けることで、文書中の単語にその説明文 / 定義文を関連付けることができる。アノテーションを利用することで、単語の意味を一意にすることが可能である[1]。さらに、アノテーションとして、統語的アノテーション（係り受け、形態素、品詞情報など）を追加することで、一般的に実用が難しい言い換えの実現も可能となる[2]。閲覧者は原文の言い換えを行うプロキシを介することによって、言い換えられた文書を観覧する。（図1）

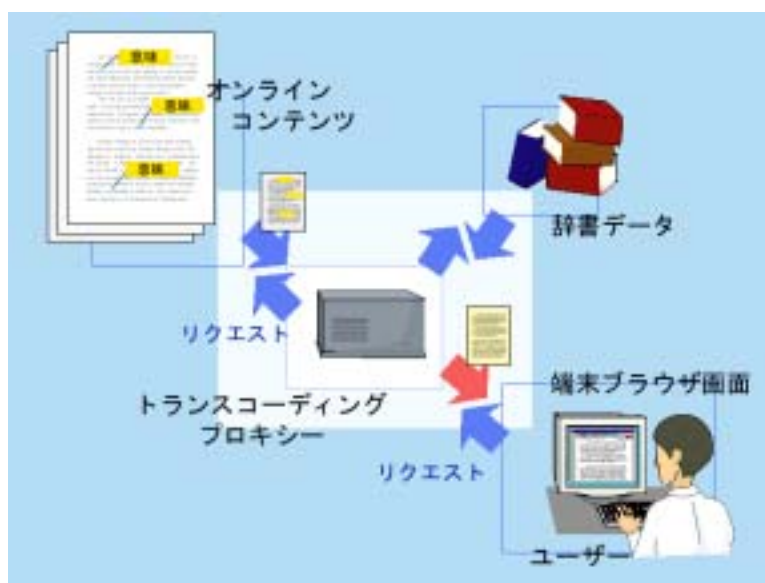


図 1
言い換え処理の流れ

言い換えは、オンライン文書の音声化に有効なほか、一般に認知的付加が少ない。これからのオンライン文書の理解促進にきわめて有効な手段になり得る。

2 . インタラクティブパラフレーズ

インタラクティブパラフレーズとは、文書全体の言い換えとは違い、文書中の単語の言い換えをユーザの選択というインタラクションに従って行えるようにするセレクトティブなパラフレーズと、言い換え後の単語についても言い換えを可能にする、インクリメンタルなパラフレーズを指す。

インタラクティブパラフレーズの種類として(1)クリックパラフレーズ、(2)リストパラフレーズ、(3)リージョンパラフレーズの3つが考えられる。

以下それぞれについて説明する。

クリックパラフレーズ

マウスで分からない単語エレメントをユーザがクリックすると、その単語の類義語または定義文を利用して言い換える。分からない語をクリックするという自然な動作で言い換えを行うことができる。(図2、図3)

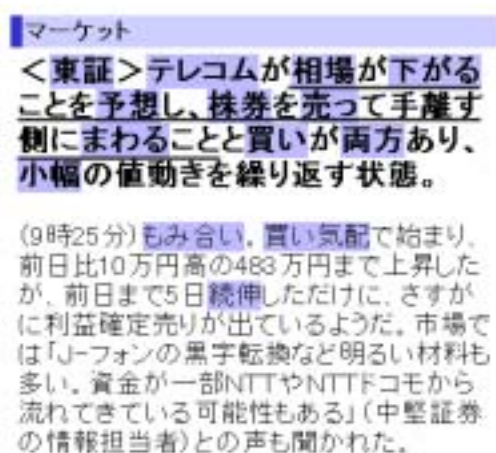
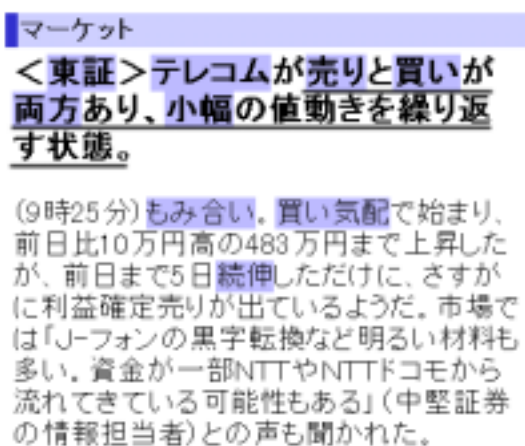


図2 クリックパラフレーズ(言い換え前)

図3 クリックパラフレーズ(言い換え後)

リストパラフレーズ

マウスで分からない単語エレメントをユーザがクリックすると言い換え可能な単語の一覧がリスト表示され、ユーザがそのうちの一つを選択するとその候補を利用して言い換える。言い換え語の候補があらかじめ一覧できるので、繰り返し言い換えする手間を省くことができる。(図4)

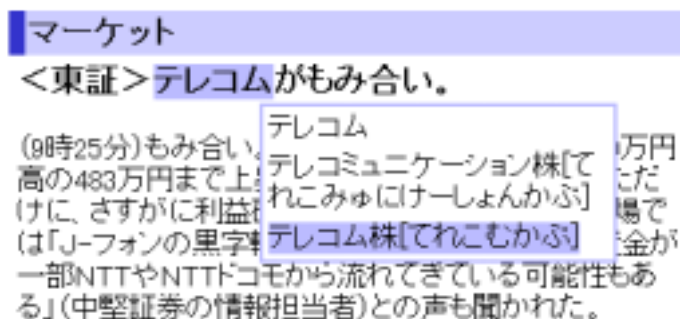


図4 リストパラフレーズ

リージョンパラフレーズ

ユーザがマウスのドラッグ操作により、文書内の領域を選択し、それに含まれるすべての単語に対しユーザ履歴を参照して言い換える単語を選別し、言い換えを行う。一度に複数の言い換え候補を選択することができるため、ユーザの手間を減らすことができる。(図5、図6)

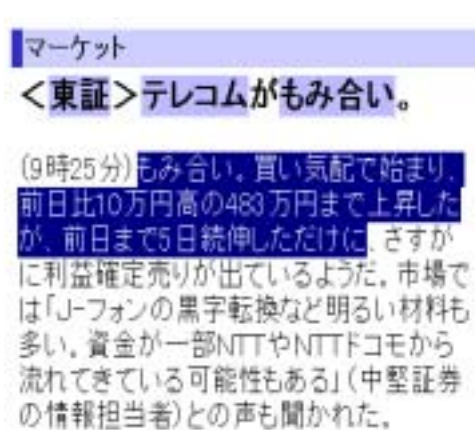


図5 リージョンパラフレーズ(実行前)

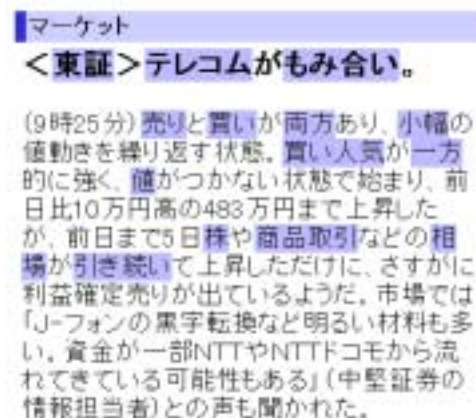


図6 リージョンパラフレーズ(実行後)

3. おわりに

説明したように、文書と辞書に統語アノテーションと語義アノテーションを付与し、それらを本発明による言い換えシステムを介し文書を閲覧することで、ユーザの理解困難な単語を含む文書であっても、ユーザとのインタラクションにより、より分かりやすく提示することができる。今後オンラインドキュメントがますます増加していく中で、本発明により、背景知識を持たない人であっても理解できる文書が増え、オンラインドキュメントがより効率的に利用されるようになると思われる。

参考文献

- [1] Katashi Nagao et al.
Semantic Transcoding: Making the World Wide Web more understandable and usable with external annotations.
TRL Research Report RT0386. IBM Tokyo Research Laboratory, 2000.
- [2] 東中竜一郎, 長尾確. アノテーションに基づく知的文書変換.
情報処理学会研究報告 2000-ICS-120 pp.33-40, 2000.