ネットニュースグループ fj.wanted のダイジェスト自動生成

佐藤　理史[†]　佐藤　円[†]

本稿では，fj.wanted のダイジェストの自動生成を実現する方法について述べる．その中心技術は，ニュース記事からのサマリ抽出法である．この方法は，言わば「斜め読みを模擬した処理」であり，まず，表層的な表現を手がかりとして，42 の特徴を抽出し，それらの特徴を用いて，記事のサマリ（カテゴリとサマリ文）を抽出する．ブラインドデータに対する実験において，本方法は，カテゴリ判定正解率 81%，サマリ文抽出正解率 76%という値を示した．抽出されたサマリはカテゴリ毎に整理され，HTML 形式のダイジェストとして出力される．このとき，元の記事へのポインタは，ハイパーテキストのリンクとして埋め込まれる．作成されたダイジェストは，WWW のクライアントプログラムによって読むことができる．

キーワード:　ダイジェスト自動生成，ネットニュース，サマリ抽出，ハイパーテキスト，ワールドワイドウェブ

Automatic Digesting of Network Newsgroup fj.wanted

Satoshi Sato[†] and Madoka Sato[†]

This paper proposes an automatic digesting system for a Network Newsgroup "fj.wanted". The key component of the system is the *summary extraction* from NetNews articles. In the summary extraction, 42 features are detected by using surface language expression patterns. Using these features, the system determines the *category* of the article and extracts the most important sentence (*summary sentence*) from the article. A blind test demonstrates that the accuracy of the category detection is 81% and the accuracy of the summary sentence extraction is 76%. The system creates the digest in HyperText Markup Language from the extracted summary sentences. The digest can be accessed via World-Wide Web.

KeyWords:　*Automatic Digesting, NetNews, Summary Extraction, HyperText, World-Wide Web*

† 北陸先端科学技術大学院大学 情報科学研究科, School of Information Science, Japan Advanced Institute of Science and Technology