# WWW用機械翻訳サービスの高速処理の研究

3T－2

山本　秀樹+　村田稔樹+　福島直士*　本多英明*

+沖電気工業株式会社 研究開発本部 関西総合研究所　*沖ソフトウェア株式会社 中部支社

## 1．はじめに

インターネットの普及に伴って、様々なネットワークサービスが生まれてきている。それらのサービスを提供するアプリケーションは今後ますます高機能になっていく。そのような世界では高速な情報処理が求められる高度なアプリケーションを低速回線や既存の端末でもより簡便に利用可能にするネットワーク制御技術の確立を図ることが必要である。

本研究では、大量の情報を必要とするアプリケーションの一つとして自動翻訳を対象に、ネットワーク側でのデータ処理を効率化し、端末への伝達速度を高めることにより、利用者の様々な要求に対する応答性を向上させる方式を開発した。

## 2. 高度アプリケーションの高速化方式

### 2.1 システムの概要

WWW ページの自動翻訳を行なうアプリケーション（WWW 用機械翻訳システム[1]）を含んだ全体システムの構成を図１に示す。WWW 用機械翻訳システムは、WWW ページを実際に翻訳する翻訳サーバと、翻訳サーバの翻訳処理結果やユーザの特性情報を管理する管理サーバから構成される。１台の管理サーバは複数の翻訳サーバを管理することができる。また、管理サーバどうしが管理情報を交換することで、より大規模なサービスを提供できるように構成している。

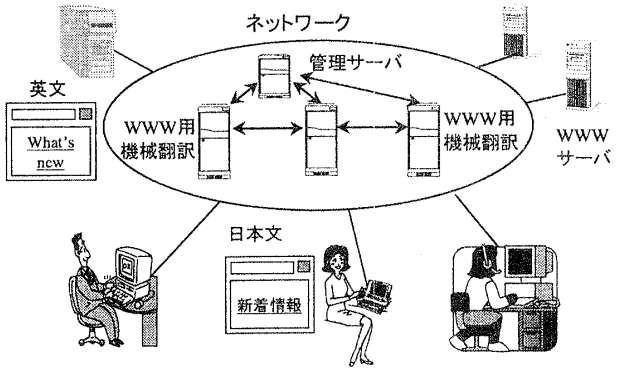


図１　WWW 翻訳環境

### 2.2 先行データ処理の最適制御

WWW 翻訳環境を利用しているユーザの挙動は、ネットワークの込み具合、翻訳対象である WWW ページの内容の複雑さ、WWW ページをすぐ翻訳する必要性があるかなどによって変化する。文献[1]の研究では、ユーザの挙動を考慮せずに常に先行処理を実行していたために余分な負荷がかかっていた。ユーザの挙動とは、翻訳ボタンを押すまでの時間と翻訳が終わっていなかった場合に再度翻訳結果を要求するかをさす。このようなユーザ特性を考慮し次のような翻訳スケジューラを開発した。

(1) 翻訳スケジューラはユーザが WWW ページをアクセスしたときに、最大翻訳時間を割り当てた翻訳スレッドを起動する。

(2) 最大翻訳時間が経過した時点で翻訳が終了していないときは (a) 同時実行可能翻訳スレッド数よりも実行されている翻訳スレッドが少なければ最大翻訳時間を短くして翻訳スレッドを継続する, (b)そうでない場合は翻訳スレッドを中断する。

実際に方式を試作し、WWW の 40 ページ(3745 文)に対するアクセスを調べた。翻訳不要ページの文数は 1161 文であり、その中で、(2)(a)で継続したスレッドが翻訳した文数を除いた事前翻訳処理した文数は 310 文 (27%) であった。従来方式では翻訳不要ページでも 100%翻訳してしまっていたが、本方式により、利用されない翻訳結果の割合を 27%に抑えることができた。また、開発した方式では、2(a)の方式により、サーバの負荷が低い場合にはサーバを有効活用することができている。

### 2.3 処理結果の蓄積単位の最適化

翻訳対象となると想定される WWW を調査したところ、企業や個人の Web サーバのトップページでは、ページの更新は頻繁に行われているが、内容が大幅に代わることは少なく、最新情報へのリンクの追加に伴い、不要となった情報へのリンクの削除といった更新が多い。また、ニュースのページは、毎日あるいは１週間単位で記事が追加される形で変更されている。従来方式［1］では、元のページのファイルの更新日付が新しくなると、翻訳結果のキャッシュを破棄して新たに翻訳を行なっている。

---

A Study of high speed WWW machine translation system

Hideki Yamamoto+, Toshiki Murata+, Tadashi Fukushima*, Hideaki Honda*

+ Kansai Laboratories, R&D Group,Oki Electric Ind. Co.,Ltd.,

Crystal Tower, 2-27 Shiromi 1-Chome, Chuo-ku, Osaka 540-6025,Japan. Hyama@kansai.oki.co.jp

* Oki Software Corporation.

そのため、これらのページに対しては部分的には内容が重複しているにも関わらず、毎回翻訳をやりなおすことになっている。

本研究では、単文単位に翻訳結果を蓄積したファイル（前回翻訳テーブル）を翻訳サーバが持つようにした（図 2）。ニュースサイトとオンラインソフトウェアサイトあわせて 10 ページ 348 文を対象に評価を行なったところ、従来方式ではページの更新時間が新しくなったため再度翻訳されてしまい再利用率は 0%であったが、本方式では約 70%（=244/348）の翻訳結果が再利用された。この実験では利用率を約 70%向上することができ、ページ全体の翻訳時間を短縮することができる。

### 2.4 ネットワーク上のサーバ間での処理結果の共有プロトコル

ネットワーク上の複数の翻訳サーバが互いに処理結果を共有するための方式として、対等の関係にある翻訳サーバ同士が互いに翻訳結果の有無を問合わすピアツーピア方式と、管理用サーバを用意しそれに翻訳サーバが問合わすことによって翻訳結果の場所を知るというサーバクライアント方式が考えられる。本研究では、翻訳結果を得るための通信コストが少ないことを重要視し、サーバクライアント方式を採用することにした。管理サーバには翻訳結果すべてを保持するのではなく、翻訳結果とそれを保持している翻訳サーバの情報のみを持つようにした。45 個のリンクを持つページをある端末と翻訳サーバの組では上から、別の組では下から順番に翻訳するという実験を行ない、各サーバ平均でほぼ 50%の翻訳結果が共有され再利用されることを確認した。

## 3. 今後の課題

### 3.1 大規模なネットワークでの各種端末を用いた実験

端末やネットワークの状況は刻一刻と変化している。単にネットワーク速度が向上して端末が高度化するといった単純な変化ではなく、これまで予想していなかった家電の端末化といったように、低速ネットワーク上の低機能な端末も増加すると予想される。そのような端末からアクセスするユーザ特性モデルはこれまでとは違ったものになるだろう。したがって，今後は本研究で提案したユーザ特性モデルの評価と改良のために、より大規模なネットワーク環境で、様々な端末を使っての実証研究が必要である。

### 3.2 コンテントによるユーザ特性モデルの精緻化

本研究では，利用者がどのような速度で翻訳要求を出しているかについてを考慮し，ユーザ特性モデルを作成した。しかしながら、同じ利用者でも見る内容や目的によって WWW にアクセスする挙動は変化すると予想される。利用者がどのような目的で WWW にアクセスしているかについてを考慮して、ネットワーク側のサービスが先行処理を制御できるようになれば、さらに応答速度を向上できるようになると考えられる。

### 3.3 他のサービスとの融合した場合のユーザ特性モデル

本研究では，翻訳サービスの利用者は人間であると仮定したが、今後ネットワーク上の高度アプリケーションが増えてくるに従い、利用者が他のネットワークサービス、すなわちソフトウェアエージェント（以下、エージェントと呼ぶ）になることが予想される。エージェントからの処理要求は、厳密には個々のエージェント毎に異なるが、ある程度は類型化できると考えられる。類型化（グループ化）した場合に、個々のグループに対してそれぞれ適切な先行処理が必要になるだろう。エージェントを含む利用者特性のモデルを考える必要がある。

## 4. まとめ

本研究の成果によって、従来から使用されている端末や、携帯性を重視した小型端末上ででも、高度アプリケーションの一つである翻訳を高速に利用できるようになる。情報ネットワークの普及に伴い翻訳などの高度アプリケーションの需要は益々増えてくる。この研究の成果は、ネットワーク上でコンテンツを加工するようなサービスに対して応用可能である。

今回の研究成果の評価を通じて新たな課題が明らかになり、さらなる高速化への手がかりを得られた。今後は、異種ネットワークにつながった大規模な環境での方式の検証や最新翻訳技術の融合[3]が必要であると考えている。

本研究は、通信・放送機構の委託研究として行われた。

[参考文献]

[1] 村田、山本、永田：「WWW のリアルタイム翻訳を可能とする機械翻訳システム」，電子情報通信学会論文誌 Vol.J79-B-I No.5 (1996)

[2] 沖電気工業:「100% JavaTM で記述された機械翻訳システム」，http://www.oki.co.jp/OKI/RDG/JIS/java/pensee/pr971217.html

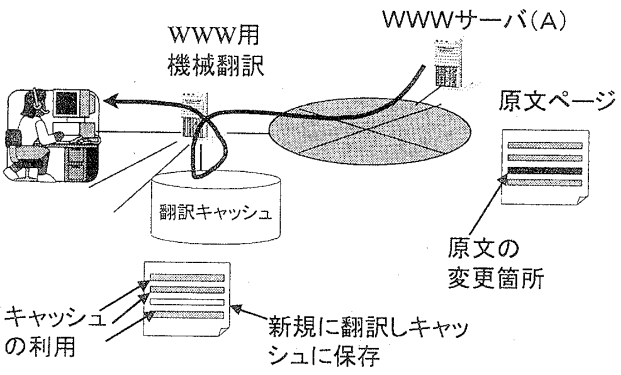


図 2　処理結果の蓄積単位の最適化