（報道発表資料）

平成14年12月3日

日本電信電話株式会社
株式会社エヌ・ティ・ティ エックス

国内の全Webページを網羅する
「新鮮情報検索エンジン」の実証実験を開始

―超高速情報収集技術をベースに最速15分前に更新された情報の検索を実現―

日本電信電話株式会社（以下NTT、本社：東京都千代田区．代表取締役社長：和田紀夫）は、ＮＴＴサイバーソリューション研究所が新たに開発した「新鮮情報検索エンジン」を“次世代のインターネット検索サービス”に適用させるための実証実験を、株式会社エヌ・ティ・ティ エックス（以下NTT-X、本社：東京都千代田区．代表取締役社長：中嶋孝夫）の協力を得て、インターネットポータル「ｇｏｏ」(*1)において、平成14年12月４日より約４ヶ月間実施します。

本検索エンジンは、8000万ページと言われる日本国内の全Webページをカバーしつつ、最短で15分前にWebサーバに掲載された情報の検索を可能にし、日々刻々と変化する各種ニュースは勿論のこと、新製品情報やスポーツ速報、最新のイベント情報などあらゆる情報を、どこよりも早く検索することを可能とします。

○狙い

インターネット上のWebページには、日々刻々と多くの新しい情報が掲載、更新されています。しかしながら、従来のインターネット検索サービスで検索できるWebページの内容は、数日から数週間前の情報であるのが実情です。また、最新情報を検索できるいくつかの検索サービスは、限られたサイトを対象に実現しているもので、両社が「goo」において、本年８月７日より提供している「最速ニュース記事検索実験サービス」（*2）も特定のニュースサイトのみを対象としています。これは、インターネット上に掲載されている全ての情報を検索対象データ化するための処理に時間を必要とするという、現状の商用検索エンジンの性能に起因します。対象データの更新頻度と、検索対象とするページの規模（多さ）とを両立させることは、大きな課題でありました。

常時接続環境により、インターネットを情報収集手段としてより活発に利用する時代に、爆発的に増加するインターネット全体を対象としながらも、

最新情報を得ることができる“次世代のインターネット検索サービス”の実現が待望されていました。

これらの状況を鑑み、NTTでは、対象を限定せずにいつでも調べたい最新情報を利用できる情報検索サービスを目指し、「新鮮情報検索エンジン」の開発を進めてきました。本検索エンジンは、日本国内のWebページ全体を１日に１度以上収集できる「超高速な情報収集技術」をベースにしたもので、これまで国内の検索サイトにおいて主流であった国外検索エンジンに対抗した、日本発の「新鮮情報検索エンジン」であります。

○**実験の概要と目的**

両社は、ブロードバンド時代のインターネット検索サービスの実現に向け、NTTが開発した「新鮮情報検索エンジン」の有効性、有用性を実証するため、NTT-Xが運用するインターネットポータル「goo」のトップページおよび「Webページ検索サービス」の検索結果ページから「最新Web検索　実験サービス」へのリンクを貼り、「goo」ユーザの皆さまへ公開いたします。

NTTは、今回開発した「新鮮情報検索エンジン」を次世代型検索エンジンと位置づけ、実環境によるシステム性能の検証を行います。また、NTT-Xは、「goo」のより一層の充実・強化を視野に，ユーザニーズの評価・検証を実施し、ブロードバンド時代のインターネットユーザの皆さまに求められているインターネット検索サービスの方向性を探ります。

○**技術のポイント**

本「新鮮情報検索エンジン」は、更新情報を高速に効率よく収集するために以下の２つの技術を用いています。これにより、１億ページ以上のWebページ情報が１日で収集・検索可能となります。

１．Web空間自動学習による超多重収集制御技術（別紙１）

バーチャルドメイン(*3)、ミラーサーバ(*4)などのWeb空間の構造を学習、判定することにより、複数の収集ロボットを効率良く制御し、同一のドメイン、IPアドレス(*5)へのアクセスが重複しないようにするなどインテリジェントな多重収集制御を実現した技術です。これは、従来の収集速度の約２倍に相当するとともに数億ページ規模まで拡張が可能です。

２．更新ページ学習収集制御技術（別紙２）

Webページの本文の内容が更新されているのか判別することにより、本文が更新されたページに関してだけインデックス(*6)変更を行なう更新ページ学習収集制御を実現した技術です。

また、このように高速に収集されたWebページを即座に検索結果に反映するためには、インデックスを高速に更新することが必要であり、上記技術に加え、今回の実験でも「最速ニュース記事検索実験サービス」と同様に以下の技術を用いています。

３．圧縮付リアルタイムインデクシング技術

各収集ページのキーワードを高速に抽出し、インデックスをリアルタイムに書き換えるとともに、インデックステーブルの圧縮を行い、転送量の軽減を図ることを実現した技術です。

○**今後の予定**

常時接続が一般化したブロードバンド時代におけるポータルサイトの更なる充実を目的に、NTTでは、インターネット検索サービスの更なる高速化、高精度化技術の開発を進めます。またNTT-Xでは、本実験で得られたデータを基に、本エンジンの「goo」への導入を目指しております。

《用語解説》

(*1)【 goo 】http://www.goo.ne.jp/

約1,300万人のユニークユーザ(下記注)を有するポータルサイト。“インターネット検索サービス”をはじめとする多彩な「検索サービス」を核に、約300万会員を有する「コミュニティ」、ニュース等の「コンテンツ」などを提供しています。

(注)日本リサーチセンターのWWW視聴率調査レポートによる視聴率などをもとに算出

(*2)【 ニュース記事検索 実験サービス 】

NTTサイバーソリューション研究所が開発した収集したWebページを即座に検索結果に反映する「圧縮付きリアルタイムインデクシング技術」の評価・検証を目的に、平成14年8月7日より「goo」において行っている公開実験。主にニュース系のWebサイト等から日々刻々と発信される情報を検索対象としており、新鮮な情報をすばやく、そして効率よく集めることが可能になっています。

(*3)【 バーチャルドメイン（virtual domain） 】

１台のサーバに複数のドメイン名を割り当て、同時に複数のサービスを異なるドメイン名で提供することです。例えば、レンタルサーバなどで利用され、１台のサーバを複数の利用者が、あたかも別のサーバとして公開することができます。なお、ドメイン名とは、インターネット上に存在するサーバなどにつけられる識別子であり、インターネット上の住所のようなものです。数字の羅列であるIPアドレスは人間にとって扱いにくいため、アルファベット、数字と一部の記号で表現されます。

(*4)【 ミラーサーバ（mirror server） 】

特定サーバへのアクセスが集中することによる負荷を軽減するために設置される、それと同じ機能を持たせた別のサーバのことです。例えば、インターネット上で、あるサーバが人気の高いコンテンツを公開している場合、そのサーバに対してアクセスが集中しサーバの処理能力を超え、アクセス不可能な状態に陥る場合があります。このような現象を避けるため、ミラーサーバを配置することで負荷分散を図ります。

(*5)【 IPアドレス（Internet Protocol Address） 】
インターネットなどのIPネットワークに接続されたコンピュータやサーバなど、1台1台に割り振られる識別番号のことです。

(*6)【 インデクシング 】
検索を高速に行うためのインデックス(索引)を作成する処理。収集した Webページから検索のキーワードとなる単語を抽出して、各キーワードとそれぞれが出現するWebページの対応を記録したデータベースを作成します。検索エンジンは、これを参照することによって高速な検索処理を実現しています。

・（別紙１）Web空間自動学習を用いた超多重収集制御技術
・（別紙２）更新ページ学習収集制御技術

<お問い合わせ先>
日本電信電話株式会社
サイバーコミュニケーション総合研究所
企画部　広報担当　落合・山下・萩野
Tel：0468-59-2032
E-mail：ckoho@lab.ntt.co.jp

株式会社エヌ・ティ・ティ エックス
広報室　鈴木・田畑・栗山
Tel：03-5224-5500
E-mail：pr@nttx.co.jp

NTT NEWS RELEASE