

プレスリリース

Web ページ記述のための世界共通語

平成 18 年 11 月 24 日

特定非営利活動法人セマンティック・
コンピューティング研究開発機構

Web ページを記述するための世界共通語 CWL(Common Web Language)を国際標準にするための活動の第1ステップとして、WWW コンソーシアム(W3C)のインキュベータ活動を開始した。CWLは、現在の Web 環境から言語障壁を取り除き、かつ Web ページのコンテンツの知的利用を可能にするための言語であり、Web の世界に新たなイノベーションをもたらすものである。この W3C での活動により、1 年以内に CWL の仕様とパイロットプラットフォームを公開し、引き続き、日本主導による国際標準化のための標準化トラックに移行する予定である。

現状の Web ページには 2 つの大きな課題がある。その一つは Web ページにおいては英語が主流になっていることである。各国のサイトでは自国語のページが多く、非英語圏では英語ページとの間に言語障壁が生じている。Google や Yahoo ではこれを機械翻訳サービスで解消しようとしているが、翻訳精度に多くの問題があり、かつ主要言語以外の言語はサービスの対象となっていない。第 2 の課題は、コンピュータが Web ページの内容を知的利用することの難しさである。現在 Web ページの記述は HTML でタグ付けされているがコンピュータがテキストの内容を知的利用するためには、このタグ付けでは不十分で、どうしても自然言語を理解することが必要になる。

W3C ではページのタグセットを特定の応用に適したものに設定できるようにしたセマンティック Web 言語 RDF/OWL を標準化してはいるが、自然言語で記述された内容をコンピュータ理解可能にするための具体的な言語システムは未だ存在していない。CWL は以上の言語障壁とページの内容理解の 2 つの課題を解決するために設計された人工言語である。

CWL で蓄積された Web ページは多言語環境で自由に利用でき、また CWL で記述された Web ページがコンピュータ理解可能であるということからコンテンツの知的利用が可能になり、さまざまな Web 情報サービスが期待できる。CWL で記述された Web コンテンツが今後国際協調の下で蓄積されていくことにより、CWL は Web グローバリゼーションの拡大と Web インテリジェンスの向上にとってエポックメイキングな言語になるであろう。

CWL は総務省 SCOPE プログラムから委託を受けて進めている「CDL の言語仕様の策定と標準化」の研究(代表 石塚満東大教授)において、国連大学で開発され UNDL 財団(理事長 Tarsicio Della Senta)が推進している UNL(ユニバーサルネットワークング言語)を、Web での

利用環境と、マルチメディア情報を中心としたさまざまな情報表現との統合を考慮して再設計したものである。CWLはUNLシステムのもつ各国語との相互変換機能によりWeb上に多言語環境を実現する。

このW3Cでのインキュベータ活動はISeC(特定非営利活動法人セマンティック・コンピューティング研究開発機構)(理事長 相磯 秀夫)、独立行政法人産業技術総合研究所(理事長 吉川 弘之)、および(株)ジャストシステム(代表取締役社長 浮川 和宣)からの提案によるものである。第一回会合は2006年11月28日に東京台場のホテル日航東京にて開催される。

【用語解説】

WWW コンソーシアム:

World Wide Web Consortium(略称 W3C)はWebの標準化を主たる活動とする国際的な組織で、現在世界中の会員機関 400 で構成されている。ディレクターはWebの開発者 Tim Berners-Lee である。米国 MIT コンピュータサイエンス人工知能研究所、フランスに本部を置く欧州情報処理数学研究コンソーシアム(ERCIM)、慶應義塾大学が世界の3拠点になって運営されている。2006年11月28日にW3C アジアホスト開設 10周年記念祝賀式典が東京台場のホテル日航東京にて開催される。

インキュベータ活動:

2006年2月にW3Cが制定したもので、萌芽的なWeb関連技術の開発促進を目的としている。標準化作業を行う Recommendation Track の前段階に位置づけられる活動。Web技術における革新性と基盤性を持っていること及びWebアプリケーションの新しい芽を育てることが条件である。CWLは5番目の活動に当たる。

CDL:

Concept Description Language とは、ISeCが推進するセマンティック・コンピューティングの基盤言語である。コンテンツの概念構造を記述することを目的とし、自然言語に限らず、多様なメディアによる情報を共通の言語モデルで表現することを可能にする。CWLはCDLの自然言語に共通な枠組みを設定したCDL.nlにUNLの語彙を付与したものである。

SCOPE:

戦略的情報通信研究開発推進制度(Strategic Information and Communications R&D Promotion Programme)は、情報通信技術(ICT)分野の研究開発における競争的研究資金制度である。ICT分野のイノベーションを生み出すことを目指し、総務省が定めた戦略的な重点研究開発目標を実現するための独創性・新規性に富む研究開発を支援する制度。今回のCWLはSCOPEの「国際技術獲得型研究開発」に分類されるテーマ「CDL(Concept Description Language)の仕様策定と標準化」の成果に基づいている。

UNDL 財団:

UNDL財団は、国連大学が2001年1月に、UNLプログラムの開発管理を行う目的で設立した非営利国際機関であり、スイスのジュネーブに本拠をもつ。国連からUNLプログラムを実施する権限を受け継いで、UNLの開発、応用、普及を行っている。<http://www.undl.org/>

UNL (Universal Networking Language):

UNL は国連の中で誕生し、国連大学高等研究所(UNU/IAS)で研究開発されてきたコンピュータのための言語である。UNL の所有権は国連にあり、全ての人々が自由に使えることを保証するために、UNL は国連の名前で特許が登録されている。

UNL は人工言語であり、コンピュータの世界において人間のための自然言語の機能を模写したものである。その結果、自然言語でもたらされる全ての情報や知識は UNL で表現することができる。またコンピュータは互いに UNL でコミュニケーションすることができるようになり、その結果、多言語で情報を発信したり、受け取ったり、理解したりするための言語基盤を人々に提供することができるようになる。

UNL システム:

UNLシステムはUNLと各言語間の翻訳機能をインターネット上を実現するもので、UNLで表現された情報や知識を各言語で表現したり、各言語で記述された情報を UNL での表現に変換したりする。UNLシステムは世界 20 カ国以上の自然言語にリンクして実際に動作している。

本件に関する報道機関からのお問い合わせ先:

ISeC 理事 横井 俊夫 (東京工科大学)

Tel 0426-37-2435 E-mail: yokoi@media.teu.ac.jp

産業技術総合研究所情報技術研究部門 副部門長 橋田 浩一

Tel 03-5298-4728 E-mail: hasida.k@aist.go.jp

株式会社ジャストシステム 主任研究員 大野 邦夫

Tel - E-mail: kunio_ohno@justsystem.co.jp