

CDL (Concept Description Language) の仕様策定と標準化 (053103002)

Specifications and Standardization of Concept Description Language

研究代表者

石塚 満 東京大学 大学院情報理工学系研究科

Mitsuru Ishizuka Graduate School of Information Science and Technology, The University of Tokyo

研究分担者

石崎 俊[†] 内山 清子[†] 橋田 浩一^{††} 和泉 憲明^{††}

横井 俊夫^{†††} 内田 裕士^{†††} 安原 宏^{†††} 朱 美英^{†††}

Shun Ishizaki[†] Kiyoko Uchiyama[†] Koiti Hasida^{††} Noriaki Izumi^{††}

Toshio Yokoi^{†††} Hiroshi Uchida^{†††} Hiroshi Yasuhara^{†††} Meiyong Zhu^{†††}

[†]慶応義塾大学環境情報学部 ^{††}独立行政法人産業技術総合研究所情報技術研究部門 ^{†††}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

[†]Keio University Faculty of Environment and Information Studies ^{††}National Institute of Advanced Industrial Science and Technology Information Technology Research Institute ^{†††}Institute of Semantic Computing

研究期間 平成 17 年度～平成 19 年度

概要

コンテンツの意味構造（概念構造）の記述を行うための言語 CDL (Concept Description Language) のコア部分（基本データ構造）と自然言語共通部分の最終仕様を策定し、標準化を提案、実現する。また当該 CDL 仕様を検証するために複数の観点で実証研究を行う。CDL は、我が国の主要な大学や研究機関の代表的な研究者によって進められつつある SeC (Semantic Computing) 計画の共通言語である。標準化に関しては、セマンティック Web や XML と密に連携することから、W3C (World Wide Web Consortium) のセマンティック Web 活動グループへ提案する。CDL を言語仕様に留まらずシステムとして位置づけて提案するために Common Web Language という名称で提案した。

Abstract

The research proposes specifications for a new language called CDL (Concept Description Language). The objective is to promote a new computation paradigm “Semantic Computing”. CDL is a family language which has a layered architecture. CDL.core and CDL.nl are common-base languages for the first step of the standardizing of the specifications. The standardizing organization is W3C (World Wide Web Consortium). We proposed CWL (Common Web Language) which combines triple language systems: CDL, UNL, and RDF/OWL.

1. まえがき

本テーマが 2005 年に SCOPE で採択されたとき、W3C(World Wide Web Consortium)では 2004 年 2 月に RDF と OWL の標準化が行われた後であり、Semantic Web に向けた研究が活発化し始めたときであった。Semantic Web が対象としたものは、自然言語のような論理的に複雑な構造を持つ表現ではなく、オントロジーで定義された対象と対象間の関係付けを主にしたものである。対象はコンテンツそのものというよりメタ的な情報であり、アノテーションや簡易な三つ組み表現形式であった。

Web 上の情報は自然言語が中心であり、コンピュータが処理するには不向きである。また英語が中心であり人類共通のコンテンツでもなかった。このような課題を解消する手段として CDL(Concept Description Language)という言語を介在させることで、人間同士のコミュニケーションと人間とコンピュータのコミュニケーションを増幅することを旨とする研究を提案した。

CDL はコンテンツ自体を記述するもので、コンテンツの構造を記述するための言語であり、基本的なモデル&シンタックスを提供する。コンテンツが自然言語で記述されたテキストの場合、その CDL 表現はテキストの内容記述に相当するもので、テキストのマークアップ言語として標準

化された SGML、それをより単純化した HTML や XML のハイパーテキストの系譜に対してより構造記述、内容記述に注力した言語である。

2. 研究内容及び成果

2.1 研究内容

本研究の CDL は、図 1 に示すようにセマンティックコンピューティングという計算パラダイムの共通言語となる。今回はその基盤になる CDL.core 及び CDL.nl の仕様を策定し、その応用のための評価実験を行った。国際標準には、W3C への提案を目指して作業した。

CDL.core の仕様は、Model&Syntax に限定して策定した。実体概念と関係概念の有向ハイパーグラフからなるもので木構造の XML や三つ組みの RDF 構文を包含するより一般的な言語を設定している。また、概念を定義するための CDD(Concept Definition Dictionary)を導入して XML 系列のスキーマや name スペース機能をカバーしている。CDL.nl に関しては、自然言語共通なモデルとするため文及び文章の構造を概念的に表現する語彙やトップオントロジーが必要となる。そのため、国連大学で開発された UNL との連携を図った。CDL の実証評価としては、概念レベルの高次検索、概念レベルの特許文書検索、概念レベルのオーサリングシステム等を実施した。

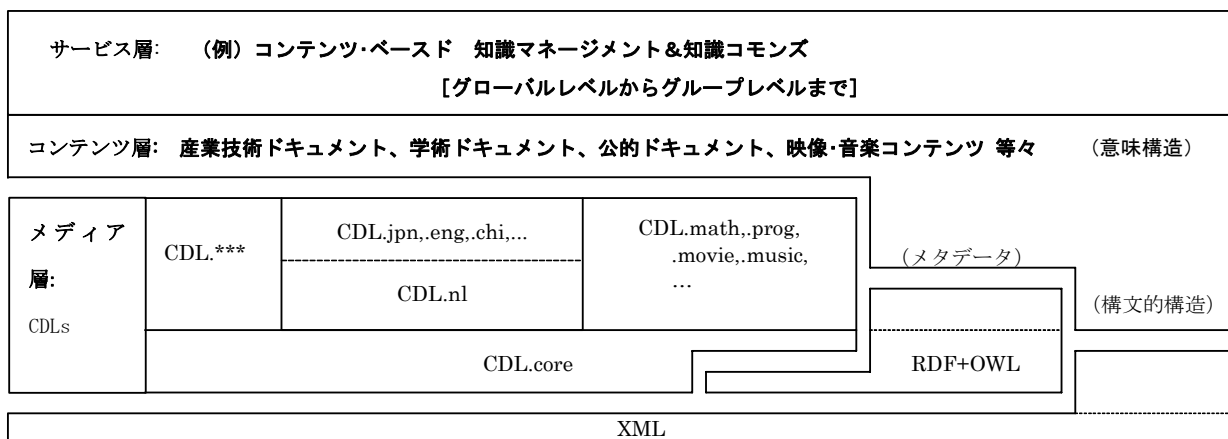


図 1. CDL の階層構造

標準化活動は W3C 内で CWL(Common Web Language)インキュベータグループ CWL XG を立ち上げた。CWL は図 2 に示すように CDL を含めた Web に絡む 3 つの言語系を統合する枠組みである。日本語や英語の自然言語は CWL.unl、リッチコンテンツ対応は CWL.cdl、XML、RDF 系統は CWL.rdf でそれぞれ分担することで情報の効率的な交換を行う。CWL.cdl の語彙知識は CWL.unl のものを利用する。また、RDF/OWL にも変換する。

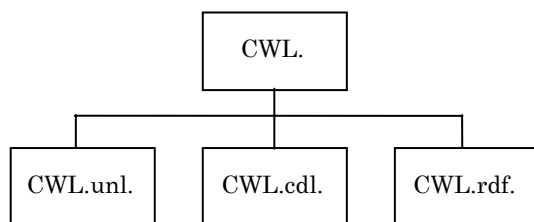


図 2. CWL の構成

2. 2 成果

CDL.core 及び CDL.nl の仕様書第 1 版を策定した。W3C CWL XG で 1 年間作業し、その最終報告書としてシステム構成及び言語仕様からなる Common Web Language が W3G のホームページに掲載されている。図 3 は本年 4 月 15 日に W3C から報道された CWL XG の News である。

► Incubator Group Report: Common Web Language

2008-04-15: The [Common Web Language Incubator Group](#) published their [final report](#). The goal of the Common Web Language is to allow the exchange of information through the Web and also for enabling computers to process information semantically. CWL allows people to describe contents and meta-data of Web pages written in natural language; the language seeks to lower language barriers and to facilitate the automatic extraction of information from Web pages. This publication is part of the [Incubator Activity](#), a forum where W3C Members can innovate and experiment. This work is not on the W3C standards track.

図 3. W3C での報道

3. むすび

言語仕様の標準化というそれ単独では利用者の関心をひきつけるのは難しい研究であったが、言語障壁解消と機械理解可能という新しい仕様の Web に付与するための提

案であり、国際的に CWL(CDL)の実現に向けてアピールできたと考えている。CWL は情報爆発の時代において重要性を増すテーマである。CWL XG は W3C における標準化トラックではない。そのためにはワーキンググループを立ち上げ、2 年間の作業が必要になる。今後も、国際標準獲得を目指してより分かり易い形でシステムを提案してゆきたい。

【国際標準提案リスト】

- [1] W3C・Common Web Language Incubator Group,A Common Web Language Incubator Group Charter,2006.10.16,2006.10.26
- [2] W3C・Common Web Language Incubator Group,CDL.core Specifications Version 1.0,2007.3.30
- [3] W3C・Common Web Language Incubator Group,Common Web Language,2008.3.30

【参加国際標準会議リスト】

- [1] W3C・Technical Plenary Week 2006,Semantic Web Interest Group Meeting,Mandelieu(France),2006.3.2
- [2] W3C・Common Web Language Incubator Group・Face to Face Meeting,Tokyo,2006.11.28.
- [3] W3C・TPAC 2007,Cambridge,MA(USA),2007.11.6

【誌上発表リスト】

- [1] 横井俊夫：セマンティックコンピューティングー知的システム・知的環境の設計原理一、人工知能学会誌、Vol.21、No.6、pp.683-690、(2006.11)
- [2] 石塚満：自然言語テキストの共通的概念記述、人工知能学会誌、Vol.21、No.6、pp.691-698 (2006.11)
- [3] 橋田浩一：オントロジーと制約に基づくセマンティックプラットフォーム、人工知能学会誌、Vol.21、No.6、pp.712-717 (2006.11)

【受賞リスト】

- [1] 石塚満、人工知能学会功績賞、2007.6.21.

【報道発表リスト】

- [1] “Incubator Group to Develop Common Web Language (CWL)”、W3C ホームページ、2006.11.6
- [2] “Incubator Group Report: Common Web Language”、W3C ホームページ、2008.4.15

【本研究開発課題を掲載したホームページ】

<http://www.w3.org/2005/Incubator/cwl/>
<http://www.instsec.org/>