

# 実世界セマンティックWebに向けて 時空間メタデータによるWeb検索の拡張

あかほに しゅんいち ひらまつ かほら  
赤埴 淳一 / 平松 薫  
すぎやま あきら やなぎさわ ゆたか  
杉山 聡 / 柳沢 豊  
さとう てつじ  
佐藤 哲司

NTTコミュニケーション科学基礎研究所

ユビキタス環境における人間どうしのコミュニケーション活性化を目的に、NTTコミュニケーション科学基礎研究所では、「実世界セマンティックWeb」の研究を進めています。Web情報から抽出した時空間に関するメタデータを用いて、Web検索を拡張する方式について紹介します。

## 実世界セマンティックWebとは

インターネット上には膨大なWeb情報が蓄積され、日々増大しています。Web情報の活用は、私たちの生活に欠くことのできないものになってきてい

ます。一方で、さまざまなセンサがネットワークでつながったユビキタス環境が実現しつつあります。Web情報やセンサ情報をうまく活用することで、人と人とのより豊かなコミュニケーションが実現できると考えられます。例えば、

人と人との会話内容に応じて関連するWeb情報を表示することで、新たな話題へ会話が展開することが考えられます。NTTコミュニケーション科学基礎研究所では、ユビキタス環境において、人と人とのコミュニケーションを活性

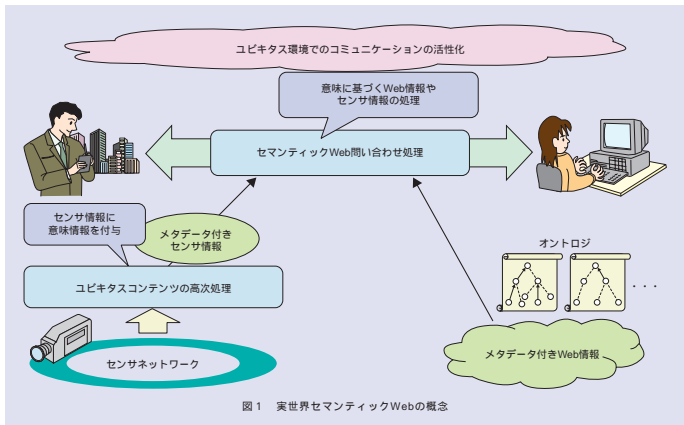


図1 実世界セマンティックWebの概念

化する技術を「実世界セマンティックWeb」と呼び、研究を進めています。

実世界セマンティックWebの概念を図1に示します。実世界セマンティックWebの実現には、大きく2つの技術が必要です。1つはメタデータの付いた情報（Web情報およびセンサ情報）を意味に基づいて処理するセマンティックWeb問い合わせ処理技術、もう1つは、センサ情報に意味情報を付与するユビキタスコンテンツ高次処理技術です。

ここでは、セマンティックWeb問い合わせ処理に関する研究として、時空間メタデータを用いたWeb検索の拡張に関する研究を中心に紹介します。

### セマンティックWeb

セマンティックWebは、WWWの創始者Tim Berners-Leeが次世代のWebとして1998年に提唱したものです。現在のWebページの大半は人が読むために作成されているため、計算機処理が困難という問題があります。例えば、「石川」というキーワードで検索した際、それが地名なのか人名なのか、あるいは企業名なのか区別がつかないため、不必要なWebページが大量に検索されてしまいます。このような場合、地名などの意味体系を高速した辞書があれば、正確かつ高次元処理ができると考えられます。

セマンティックWebでは、辞書に相当するものをオントロジと呼びます。また、オントロジに基づいて、Web情報に付与された意味記述（例えば、あるWebページの「石川」という記述は店舗名を表す）をメタデータと呼びます。

### 時空間メタデータによる検索の高度化

実世界に関する情報を扱う際には、時間情報と空間情報の扱いが重要となります。例えば、夜遅くに到着した旅先で食事をしたいという状況では、「今営業中の最寄りの飲食店」の情報が必要と考えられます。つまり、「今営業中の」という時間に関する検索条件と、「最寄りの」という空間に関する検索条件が必要です。

時計やGPS（Global Positioning System）によって、現在時刻や現在位置が取得可能ですが、このような情報をキーワードで検索するのは困難です。特に、「今営業中の」という時間に関する検索のためには、Webページから営業時間を取得し、現在時刻が含まれるかどうかを計算する必要があります。また、Webページに定休日に関する情報があれば、今日その日が定休日ではないことも調べなければなりません。

このように、時間に関する情報は、「17時から23時」という「時間区間」だけでなく、「曜日」「月日」などさまざまな意味体系が存在します。したがって、時間に関する検索には、時間に関する辞書に相当する時間オントロジが必要となります。同様に、空間に関しても、「地名」や「緯度・経度」などの空間オントロジが必要となります。

そこで、NTTコミュニケーション科学基礎研究所では時空間メタデータによるWeb検索の拡張方式を考案しました。本方式では、時間や空間に関するオントロジに基づいて、メタデータが付与されます。時空間オントロジとメタデータの例を図2に示します。図2の左側はセマンティックWebの階層構成（オントロジ層、メタデータ層、コンテンツ層）を表します、オントロジ

層から順に説明します。

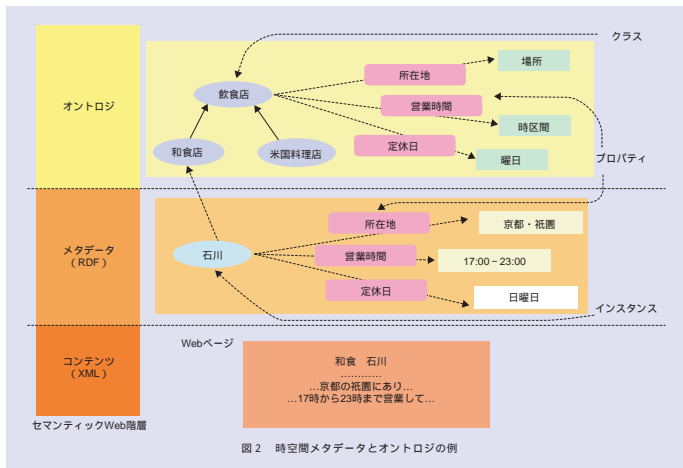
オントロジ層では、クラス間の階層関係やプロパティが定義されます。図2では、以下を表しています。

- ・「和食店」クラスおよび「米国料理店」クラスは、「飲食店」クラスの子クラス（サブクラス）である。
- ・「飲食店」クラスには、プロパティとして、「所在地」「営業時間」「定休日」がある。
- ・「所在地」プロパティの取り得る値の属するクラス（値域）は「場所」クラスである。同様に、「営業時間」および「定休日」プロパティの値域は「時間区間」および「曜日」である。

メタデータ層は、オントロジに基づいてWebページなどのコンテンツから抽出したメタデータが記述されます。図2では、コンテンツ層にあるWebページから、前述のオントロジに基づいて、以下のメタデータが記述されています。

- ・「石川」という店舗は「和食店」クラスに属する（これをインスタンスと呼ぶ）。
  - ・「石川」の「所在地」は「京都・祇園」である。同様に、「営業時間」は「17時から23時」で、「定休日」は「日曜日」である。
- セマンティックWebにおける検索式は、オントロジに基づいて記述されます。例えば、時間に関するプロパティを用いて、「今営業中の飲食店」という検索式は次のように記述できます。
- ・「定休日」が現在時刻の曜日でなく、現在時刻が「営業時間」に含まれる「飲食店」のインスタンス。

例えば、「現在時刻が」「火曜日の22時」であれば、前述の「石川」が検索結果として得られます。



### 時空間メタデータによるWeb検索システム

NTTコミュニケーション科学基礎研究所では、前述のような「時空間メタデータによるWeb検索システム」の構

築を進めています。本システムの構成を図3に示します。

本システムでは、例えば「今営業中の飲食店は？」という日本語による検索を可能としています。日本語による質問文は、質問解析モジュールにより、

時空間情報が抽出され、検索式が生成されます。この質問解析に、NTTコミュニケーション科学基礎研究所で開発した質問応答システム「SAIQA (System for Advanced Interactive Question Answering)」の質問解析

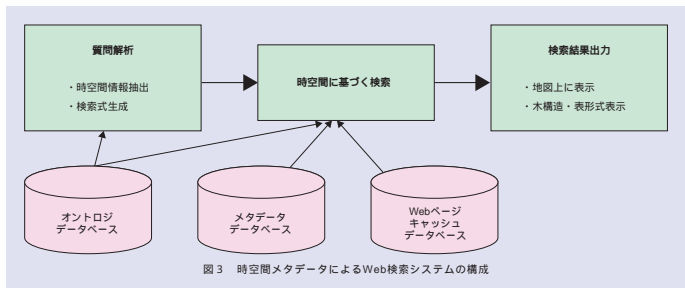




図4 時空間メタデータによるWeb検索の例

モジュールを利用しています。生成された検索式に基づいて、Webページから収集されたメタデータや、Webページのキャッシュが検索されます。検索結果は地図上や木構造、表形式で表示されます。

時空間メタデータによるWeb検索例を図4に示します。ここでは「世界文化遺産の金閣寺の近くの停留所はどこ?」という日本語文が入力として与えられます。図3の質問解析モジュールにより、「金閣寺」を検索し、さらにその近くにある「停留所」のインスタンスを検索する検索式が生成されます。検索結果は、「金閣寺」と「金閣寺前」停留所が、地図上に表示されます(図4)。また同時に、これら2つの検索結果が二次元および三次元の木構造として表示されます。

本システムと経路探索システムを組み合わせて、目的地に到着したときに利用できるサービスの検索が可能となります。また、サービス利用時間の予測と組み合わせて、次に利用すべき

サービスの検索が可能です。この際、各サービスで用いているオントロジが異なるケースに対処する必要があります。このような場合には、近似オントロジ変換手法<sup>(1)</sup>が適用可能です。

#### 今後の展開

図1に示した実世界セマンティックWebの実現には、センサ情報に意味情報を付与するユビキタスコンテンツの高度処理技術も必要です。その第一歩として、GPSなどのセンサで取得された人やモノの膨大な移動軌跡データを効率的に扱う手法の研究を進めています。例えば、イベント会場での人の移動軌跡から、同じような展示物を見て回っている人を効率的に探し出すことができます。このような同じ興味を持つ人々に対して、アドホックな電子掲示板を提供し、展示物に関連したWeb情報を提示することで、コミュニケーションの活性化が期待できます。

#### 参考文献

(1) 赤塚・平松・佐藤:「セマンティックWebの

動向と近似オントロジ変換の考察」NTT技術ジャーナル, Vol.15, No.3, pp.56-58, 2003.



(後列左から) 柳沢 豊/ 赤塚 淳一/  
佐藤 哲司  
(前列左から) 平松 薫/ 杉山 聡

膨大な情報を扱うためには、メタデータなどの意味情報を考慮した処理が不可欠です。人と人のコミュニケーションの活性化に、この膨大な情報を役立てられるように、研究を進めています。

#### 問い合わせ先

NTTコミュニケーション科学基礎研究所  
社会情報研究部  
TEL 0774-93-5234  
FAX 0774-93-5245  
E-mail akahani@cslab.kecl.ntt.co.jp