

知的 Web 情報検索システムによる電子納品データベースの有効活用に関する研究

○武蔵工業大学 学生会員 戸谷彰吾
武蔵工業大学 学生会員 田村郷司
武蔵工業大学 正会員 皆川 勝

1.はじめに

現在、XML(eXtensible Markup Language)により記述されたデータベースがWebデータベースとして利用されている。建設業界においては、国土交通省(旧建設省)が成果物の電子納品の標準化に向けてXMLを採用した。また、他分野においてもEC(Electronic Commerce)、KM(Knowledge Management)にXMLが用いられつつある。これは、インターネット上でデータを柔軟に扱えるというXMLの特性が上記の各活動に有効と判断されたことによる。

電子納品されたWebデータベースを有効活用するためには、大量のWebデータベースの中から利用者が必要とする情報を柔軟に検索する機能が不可欠である。しかし、XMLによりデータ構造は明確に定義されるものの、そのデータが持つ意味は定義されていない。そのため、データの意味を理解した検索は、現状では不可能である。したがって知的な検索を実現することによって、Webデータベースにおける知的情報を有効活用することが望まれている(図-1)。

これを可能にする技術として2001年にW3C(World Wide Web Consortium)はSemantic Webを提唱した。XMLデータでは、その構造を定義するタグが定義される。Semantic Webはタグの意味を定義するメタデータ(図-2)をそれに付加することにより、XMLデータの意味を表現する技術である。そして、その定義を検索エンジンが読み取ることによって、知的な検索が実現される。しかしながら、現時点ではSemantic Webはその概念が提唱された段階であり、実現までには至っていない。そこで、本研究ではSemantic Webで提唱された概念を既存の技術により実現するためのプロトタイプ検索システムを試作し、電子納品データに適用することにより、その有効性を検討する。

2.学科用語集について

現在、XMLベースのWebデータベースとして本学の都市基盤工学科学科用語集がある(図-3)。

システムの概要(図-4)として、XMLデータを格納するサーバに米国エクセロン社の開発したXMLデータサーバであるeXtensible Information Server3.0(以下、XIS)を使用した。XISはXMLデータをツリー構造のかたちでそのまま格納することが可能である。また、XMLパーサ、XPathプロセッサ、XSLT

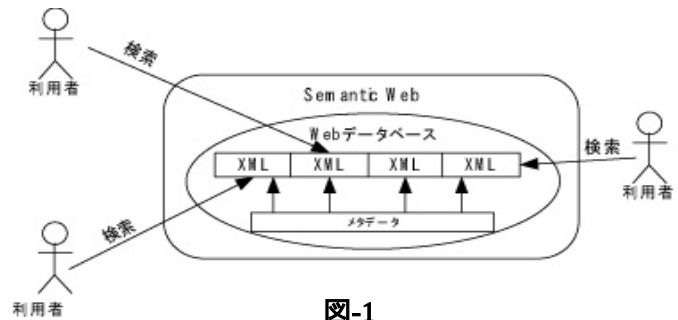


図-1

```
<rdf:Class ID="Ancestor">
</rdf:Class>

<rdf:Class ID="Parent">
  <rdf:subClassOf resource="#Ancestor">
</rdf:Class>

<rdf:Description ID="hasFather">
  <rdf:type resource="rdf:Property"/>
  <rdf:domain resource="#Ancestor"/>
</rdf:Description>
```

図-2

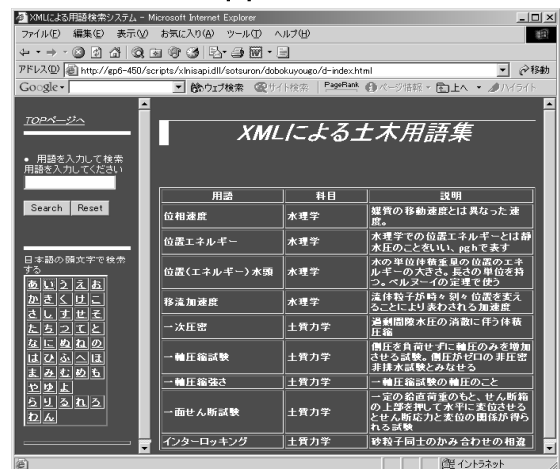


図-3

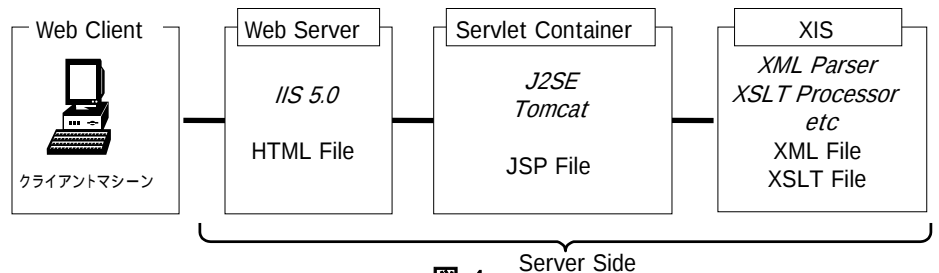


図-4

キーワード：Web、XML、Semantic Web、Thesaurus、Ontology

連絡先：武蔵工業大学工学部〒158-8557 東京都世田谷区玉堤 1-28-1 Tel&Fax03-5707-2226

プロセッサ等をサポートしている．Web クライアントであるクライアントマシンがデータにアクセスする場合は，XML データをサーバサイドで HTML に変換処理させるため，クライアント側の環境に依存することなく表示させることが可能である．JSP とはサーバサイドで処理を行う Java プログラムである．特徴として OS 環境や Web サーバ環境などの制約を受けないため，Java 環境さえ構築すれば OS などの環境に依存することなく使用することが可能である．また，CGI(Common Gateway Interface),ASP(Active Server Pages)などに比べて処理速度やサーバに対する負荷なども優れている．

3.用語集の現状と改善点について

現状の学科用語集のシステムとして、1.サーバ側で登録した用語を Web からクライアントが閲覧できる機能、2.クライアント側でクライアント本人が用語集に登録したい用語をサーバ側に登録できる機能、3.クライアントが登録した用語情報をサーバ管理者にメールで報告する機能がある．ここで、問題点として 1.クライアントが用語集から欲しい情報を得ようとするとき、もし 1 つの用語が 2 つ以上の意味を持っている場合、例えば「応用力学の安全率」と「土質力学の安全率」の意味は異なる．これは専門知識がある者であれば自分で違いを分かっているため問題無いかもしれないが、本システムは素人に分かり易く情報を提供する事を目的としているため、Thesaurus(類似語)を処理できるシステムを作る必要がある．また本システムのデータベースは各科目ごとにデータベース化してあるため、科目同士の Ontology(存在情報)を通信し合い互いの Thesaurusを確認する必要がある．2.クライアントがサーバ側に用語を登録しようとする時、もし同じ意味の単語が既に登録されている場合、2 重登録という問題が発生する．現状では登録内容がサーバ管理者にメール通知されるため、管理者、自らがサーバ側のデータベースにアクセスしてデータの修正を行うことができる．しかしこれは手動的作業であって作業の自動化が必要である．自動化するにあたって、登録される新規用語とデータベースにある用語の Ontology の判断で 2 重登録を防ぐことができる．システム概要は図-5 参照．

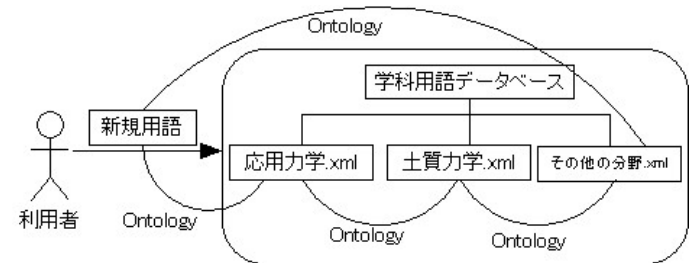


図-5

4.改善システムのモデリング

システムを構築する時、システムの設計図を作成しなければ、もしその後システムの改善を行いたいとき容易に改善を行うことが出来ない．今後既存のシステムを改善していくにあたってシステムをモデリング(設計図の作成)する必要がある．また、本システムでは JSP などの Java の技術を用いてシステムを構築しており、Java のようなオブジェクト指向言語を用いた開発にはオブジェクト指向分析設計ツールである UML(Unified Modeling Language)が採用されている．UML の設計様式で記述された設計図から Java を生成していくことが目的である．

本システムの今後の改善をモデリングした．今回は「ユースケース図(図-6)」「概念図(図-7)」を示す．

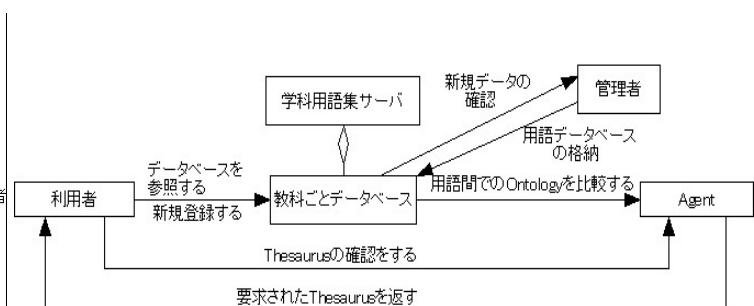
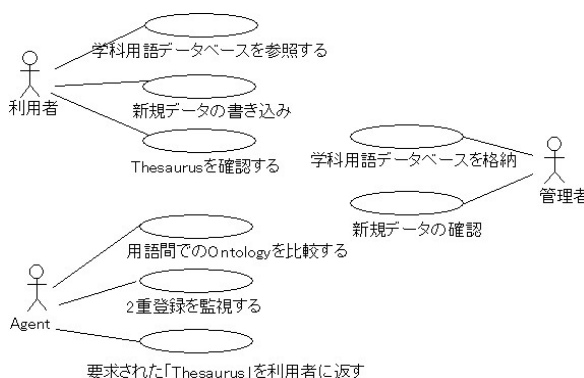


図-7

5.まとめ

XML データベースを用いた学科用語検索サイトは現在のシステム体系では土木業界に素人である利用者にとって、十分な対応が出来ていない．利用者と Agent との関連を明確にモデリングし Semantic Web を実現できるシステム体系に今後、移行していきたい．

参考文献

Henry Vayard: XML DATABASE and the SEMANTIC WEB, 2002.3
オージス総研:UML 入門教室,2002.4