

タグ情報を用いたWeb文書解析システムの 試作とその知見

—文書の構造と文書の意味に関して—

An idea for “Web Document analyzer” using markup tag.

—Is there any relationship between structure of the document
and its semantics?—

春木 良且

Yoshikatsu HARUKI

1. はじめに

「Webページは、HTMLの文書インスタンスである」。おそらく、このごく基本的、且つ最も重要な技術的前提すら、多くのWebページの作成者、管理者には理解されていないであろう。これは推測ではなく、後述のように筆者によるWebの実態調査結果からも明らかな事実であり、官公庁や大学、研究機関などの、技術的な正当性を保つべきWebページですら、そのことが考慮されていないものが多々ある。

言うまでも無く、Internet上のWebシステムは、現在においては社会の様々な局面で欠くことのできない重要なサービスとなっており、Webブラウザが旧来称されていた「キラーアプリケーション」といったレベルを超えた、いわば「情報インフラ」と言っても決して過言ではなからう。

であるがゆえに、Webに要求されるのは、アクセシビリティを高めることによって、ユーザに対するバリアフリーを実現することである。派手なデザインや充実したコンテンツなど、表面的で比較的わかりやすい側面がWebでは云々されることが多い。確かに、情報伝達手段である限りそれらも重要ではあるが、それはア

アクセシビリティが保証された上で、初めて考慮されるべき要素である。情報インフラは、ユニバーサルサービスという側面を持つということは忘れてはならないし、また情報システムの管理者は出来る限り多くのユーザに便益を提供しなければならない。

本研究では、Webのアクセシビリティの技術的な源泉であるHTML (Hyper Text Markup Language) に着目する。HTMLは、一般には、文書の装飾機能を持ったものと捉えられているが、後述のように、本来SGMLのサブセットとして、Web文書の文書構造を記述(マークアップ)する機能を持ったものとして設計されている。文書には、構造があるがゆえに装飾がなされるという前提は、マークアップ言語を理解するために重要である。

文書の構造は、すなわち文書の意味構造であり、タグ情報が文書の構造を表現するとするならば、タグを利用することによって、文書の意味情報の解析ができるものと筆者は考える。こうした仮説を元に、実験システムを作成し、Web文書解析の様々な実験を行った。以下に、それらにより得られた知見について報告する。

2. マークアップ言語とその技術経緯

2. 1. HTMLの標準化と現状のWebページの状況

特定の技術に対するアクセシビリティを高めるには、技術仕様が明確且つ公開されたものである必要がある。技術標準がデファクトなもの(事実上の標準・de-facto standard)で留まる限り、市場の変化によって技術が揺らぐ可能性があり、また開発競争のために技術仕様が特定企業内に独占される。こうした状況は、旧来の汎用機のOS技術を巡るIBMの技術戦略に見るように、特定企業の利益に資するだけであり、さらに現在のPC上の主流商用OSに見るように、技術に対する幅広い観点からの評価や修正がなされないため、ひいてはその技術そのものの発展が期待できないと

いった状況も起こりうる。そのため、公的な側面を持った団体などによって策定され公開された、規則上の標準（de-jure standard）が存在意義を持つと言える。

1990年代以降における大きな技術的トレンドは、言うまでも無く、Internetである。特にInternet上のWebシステム（World Wide Web）は、旧来類似したシステムが存在しなかったこともあり、Internetの持つネットワークの経済性の極大化といった側面と相まって、1990年代初頭から大きく注目を集めてきた。

元来、Webシステムは、1980年代末にスイスの欧州素粒子物理学研究所（CERN）で開発されたInternet上のアプリケーションである。今更述べることではないが、Webという名前の通り、ネット上に存在する様々な文書を関連づけてネットワーク化する機能を持っている。

この文書相互の関連付けを「ハイパーリンク」と呼び、ハイパーリンクによって関連付けられた文書そのものを「ハイパーテキスト」と呼ぶ。このハイパーテキストの概念自体は、1965年に社会学者であるテッド・ネルソン（Theodor Holm Nelson）の提唱による、人間の思考モデルをベースにしたものであると言われている^[1]。すなわち、Webシステムの考え方自体は、決して新奇なものではない。

Webシステムを支える技術的な要素としては、次の3点が指摘できる^[2]。

- ① 文書をマークアップ言語HTML（Hyper Text Markup Language）で記述する
- ② URL（Universal Resource Locator）表記を用いて文書の所在情報を一意的に特定する
- ③ プロトコルhttp（Hyper Text Transmission Protocol）により文書を取得する

これらは、端的に言って、Webシステムのアクセシビリティを

極大化するために機能する技術要素と言っても過言ではない。すなわち上記②は、IPプロトコルと共に、ネットワーク資源そのものへのアクセス手段として機能し、また③は文書データのやりとりに関して機能する。

文書データ記述に関する①が、情報そのもののアクセシビリティに関わるものである。②や③だけでは、最終的な情報交換は不可能である。その意味では、アプリケーションとしてのWebシステムを支える重要な技術ポイントであると言えよう。

マークアップ言語 (markup language) とは、文字列のキャラクタコード以外の様々な属性情報、特に書式にまつわる情報を、定義された記号によって、文書中に埋め込んでいく記述方式を言う。

Webページの記述言語であるHTMLは、汎用のページ記述言語のSGML (Standard Generalized Markup Language) を元に、CERN (欧州素粒子物理学研究所) で情報インフラの整備に携わっていたティム・バーナーズ・リー (Tim-Berners Lee) によって、URLの概念などとともに、基本となる仕様が作成された。

つまりHTMLは、技術的にはSGMLのサブセットという位置づけではあるが、実際問題として初期のHTMLには、DTDもなく、技術的には厳密なSGML応用言語とは言えないものであった。後にバーナーズ・リー自身も認めているが、SGMLとの関係などは曖昧であるといった指摘がされることが多い^[3など]。また、タグの設計にも本来のページ記述という観点からすれば多くの問題点がある。

Internetの一般化と共に、市場規模を拡大して行ったWebシステムにおいては、特にクライアント側のアプリケーションであるWebブラウザの製品開発競争が、1990年代半ば頃に激しく繰り広げられた。特にイリノイ大学の学生によるシステムを基にしたNetscapeと、MicrosoftによるInternetExplore (IE) の市場争奪戦

は、MSのOSとブラウザの抱き合わせ販売を巡る、独占禁止法の抵触に関する訴訟もあり、記憶に新しいところである。

このブラウザの製品競争（いわゆるブラウザ戦争）において、主要な技術的差別化ポイントとされたのが、前記①で指摘した、HTML自体の仕様拡張、すなわち独自のタグの採用であった。例えばNetscapeでは、点滅テキスト、背景イメージファイル、フレームなどがあり、IEでは、スクロールマーカー、サウンド、フォントといったタグが独自に追加されていた。但しこれらの追加されたタグの殆どは、文書の装飾機能を中心としたものであり、SGMLの本来の技術的な主旨からすれば、余り適切なものとは言えない。その点については、後述する。

HTMLは厳密な仕様に基づいて作成されたものではないということと、企業の恣意的な仕様の拡張がなされたといったことは、明らかにWebページへのアクセシビリティを低下させる。実際に最近まで、クライアント側ブラウザがNetscapeとIEの場合で切り替える構成のWebページが多かったのも、そうしたことの表れであると言える。

こうした点から、ネットワーク技術においては、標準化団体による規則上の標準の役割は決して小さくない。Web技術を含め、Internetに関わる技術の標準化団体としては、World Wide Web Consortium（W3C）があり、デファクト技術の国際的な標準化推進やプロトタイプの研究開発を行い、その成果として多くの技術仕様が公開されている。W3Cは、1994年にティム・バーナーズ・リーらが中心になって組織された国際的な学術研究開発組織であり、マサチューセッツ工科大学計算機科学研究所（MIT/LCS）、フランス国立情報処理自動化研究所（INRIA）、慶應義塾大学・SFC研究所がホストとして共同運営に当たっている。

HTMLの仕様に関しては、厳密なSGML応用言語として、文書型定義（DTD・Data Type Definition）がW3Cから公開されている。

それは、Internet上に無数に存在しているHTML文書が、そのDTDのインスタンスとして正当であること、すなわち文書インスタンスがWebクライアントのブラウザ等のUser Agent側が確実に解釈できるものであるということを保障する。これは、Webページのアクセシビリティを高めるということを意味する。

理論的に言えば、DTDは文書インスタンス毎に定義してもかまわないが、Webページの場合、実際問題として文書作成者がDTDを定義するといったことはまず考えられない。そのため、一般にはW3Cで公開されているDTDを外部DTDとして利用する。そのために文書インスタンス中では、DOCTYPE宣言（DOCTYPE Declaration）を用いて、当該文書インスタンスが基づいている外部DTDを明示するものであり、これによってブラウザ側は解釈すべきタグの範囲が確定する。例えば以下に示すDOCTYPE宣言は、W3Cで公開されているHTML 4.0のDTDを指定している。

例：

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
```

これは言わばHTMLのアクセシビリティを高めるための、最も基本となる要素である。実際にW3Cでも、「One should place a DOCTYPE declaration as the very first thing in an HTML document. Declaration is mandatory for most current markup languages and without one it is impossible to reliably validate a document.」と、かなり強い調子でDOCTYPE宣言の必要性が明記されている^[4]。

しかし、前述のように現状の多くのWebページでは、DOCTYPE宣言が省略されているものが多い。例えば筆者のサーベイでは、yahooに登録されている東京都内の大学127校のWebページのうち、トップページにDOCTYPE宣言が無いものが、実に42校にも上っている（平成17年12月現在）。神奈川県内の大学で見

てみると、その割合はさらに低下する。少なくとも大学のWebは、技術的な正当性を持ったものでなければならない。それが教育研究機関としての最低限の務めだと思うのは、筆者だけであろうか。

これらの例で見るように、一般にはHTMLの機能が正確に認識されているとは思えない。おそらくは、前述したブラウザ戦争において、HTMLに装飾機能を中心とした仕様の拡張がなされたという点が大きく影響を与えているように思える。しかし、HTMLは単なる装飾言語ではなく、SGMLのサブセットとして、文書構造を示すものとして捉えねばならない。以下に、SGMLの本来的な機能について確認する。

2. 2. SGMLの機能

SGMLが広く知られるようになったのは、1980年代初頭に、公文書フォーマットとして米国防総省のCALSSプロジェクトで採用されたのが端緒であると言えよう。1980年代に、電子技術の向上とともに軍事兵器が急速にハイテク化されて行ったのにつれ、兵器の取扱いマニュアルなどドキュメントの量が膨大なものとなって行き、それらの検索や配布、維持などが困難なものとなって行ったという事情が背後にある。CALSSは当初、こうした文書の電子化を目的としたものとして開始した。空母カールビンソンのマニュアルが、SGMLによって電子化された結果、重量23tもの空母の喫水面が5cm上がったといった、その成果を示すような噂も喧伝されている^[5]。

SGML自体は、異機種間でのドキュメント交換を実現するために、1960年代末にIBM社のCharles Goldfarbらによって開発されたGML (Generalized Markup Language) を起源としている。「マークアップ」という用語は、元々出版業界で使われていた、記号化された植字命令を原稿に挿入するということを意味する^[3]。IBMでは、1970年代からマニュアル用としてGMLの長い使用実績があ

り、マニュアル・テキストとして約900万ページ（1000ページの書籍9000冊分）のデータベースが蓄積されていると言われている^[6]。これをもとにANSI（米国標準化団体）がSGMLとして規格化し、さらにISO（国際標準化機構）が標準規格として承認している。（参照番号 ISO 8879:1986）

そのマークアップ言語の技術的な本質は、端的に言えば文書のコンテンツとそれに対する体裁情報を分離したという点にある。文書の体裁を切り離すことで、文書データそのものはプレーン・テキストとして表現できるため、異機種間でのデータの交換が可能となる。また体裁に関する情報も、テキストによる表現で記述するが、TeXのように文書の実出力レイアウトイメージそのものに着目するのではなく、文書の論理構造を記述する。文書の装飾は、文書の論理構造の表象と考えるわけである。

これにより、SGMLによる単なる異機種間のデータ交換のみならず、文書が表現している情報そのものの交換可能性を意味する。実際に、現在ではSGMLの技術的に正当な拡張形式であるXML（eXtensible Markup Language）によって、各団体や業界毎の文書フォーマットの標準化が試みられており、特に電子商取引を行う際のデータフォーマットとしても注目されている。それを前提として、各業界内などで各々の業務プロセスやデータ構造などをも統一する、ビジネスフレームワークを作る動きもあるということからも、文書の論理構造の記述が、情報そのものの表現として重要な意味を持っていることが推測される。本研究ではそこに着目するが、その詳細については次章で述べることにする。

SGML 文書は、①SGML宣言、②DTD（Document Type Definition・文書型定義）、③文書インスタンスの3つの要素から構成される。この内、SGMLの技術的な中心となるのがDTDであり、簡単に言えば文書の構造と文書の各要素にマーク付けをするための規則を定義したものである。

これら一連の技術的特徴は、HTMLにおいても変わることはない。その意味では、DTDと文書インスタンスの接続関係を示すDOCTYPE宣言は重要であるが、こうした本来の技術的な趣旨は、Webが一般化されて多くのWeb文書が生まれるに従い、形骸化して行ったというのが現在のWebを巡る技術的状况と言えよう。

3. システムの仕様とその試作

前述のように、HTMLはSGMLのサブセットとして、文書の論理構造を記述するものである。例えば、タグ<H1>は、レベル1見出し、いわゆる大項目を示す。装飾機能の側面から見ると、それは文字サイズが大きく表現されるものとして捉えられている。しかしHTML標準によれば、タグの表示レベルに関しては、基本的にクライアント側のブラウザに委ねられている。例えばボールド体を使う、下線を引くなど、大項目を文字サイズ以外で表現してもかまわない。本来のタグの意味としては、大項目であるがゆえに文字サイズが大きく表現されると解釈されねばならない。つまり、文書の表現形式は、それそのものが意味を持つのではなく、あくまでも文書の論理構造の表象にしか過ぎない。

ここで明確にしておきたいのは、文書は情報を表現する媒体であるが、さらにタグは文書を表現する手段であるということである(図1)。

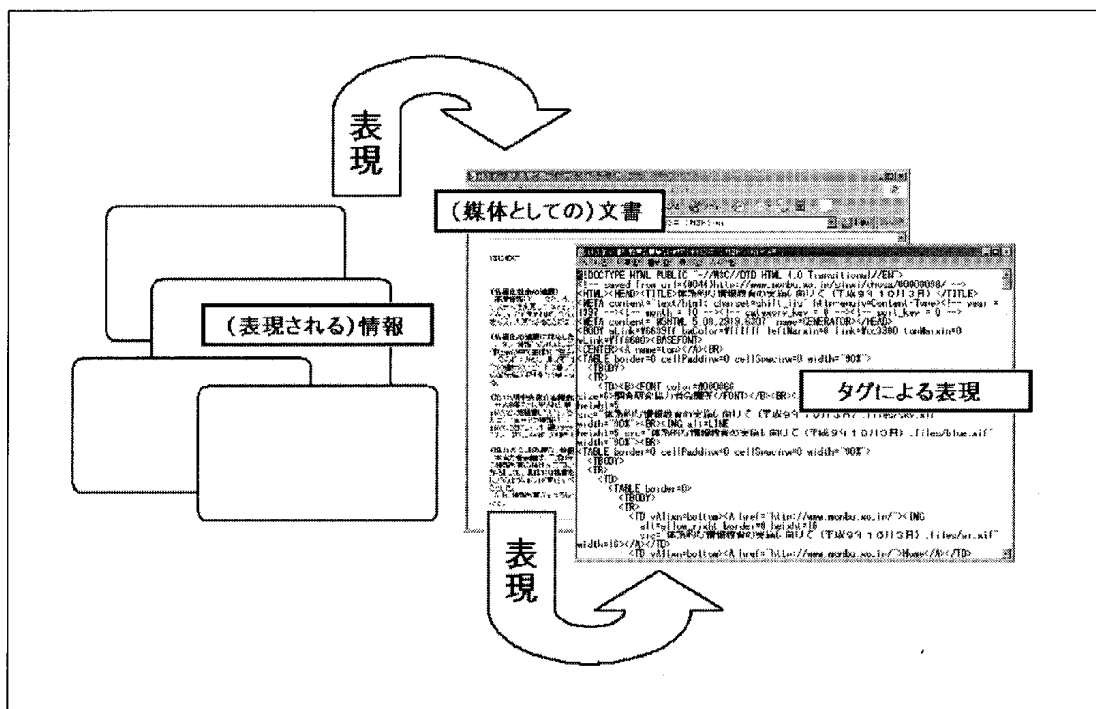


図1 表現対象と媒体の関係

であるとするならば、タグによって表現されている文書の論理構造は、その文書が表現している情報の意味を表したものであると考えることができる。つまり、文書の作者は、例えば文書によって表現されている「情報そのもの」の「ひとまとまり」に対して、文書上は「大項目」を与え、それがタグ「<H1>」で表現されているというわけである。

これを前提として、筆者らは「Web文書を構成するHTMLのタグを用いて、逆に文書が持っている情報の意味内容を得ることが可能であるはずである」と仮説付けた。すなわち、タグ情報の解析によって、文書が表現している情報の解析も可能なのではないかと考えたわけである。そしてこうした仮説を検証するために、実システムを試作し、どの程度の精度での解析が可能であるのか、HTMLタグを用いた文書の解析アルゴリズムを実験した。

具体的な検証課題としては、以下の3点を挙げた。すなわち、タグ情報を使った①文書のサマリー、②文書の内部構造、特にア

ウトライン情報の取得、③文書の外部構造、といった解析が、どの程度の精度で可能かといったことの検証である。

文書のサマライズに関しては、自動要約技術として、重要語を抽出するもの、形態素解析をするものなど、自然言語処理の文脈のものが様々に研究されている^[7]。筆者らは、文書の中で強調的な表現をされている部分に着目した。文字サイズ、文字色、太字、強調など、通常のテキストとは異なった表現をしている部分が、文書の作者が重要だと考えている部分と考えたわけであるが、実際のWebページでは、前述のようにHTMLを装飾機能として捉えているものも多く、項目を示す「<H>」をも抽出した。おそらくこうした処理と辞書、データベースなどを含めた自然言語処理を統合させることで、より精度の高いサマリーが得られるものと思われる。

文書の内部構造としては、文塊、リスト表示、表、図など文書中の情報要素（いわゆる情報チャンク）に着目した。また罫線は、文書を構造化する視覚的要素として機能していると考えた。またリンクは、アンカーとリンク先の詳細を明示することにより、ドキュメント間の意味関係が解析できるものと考えた。

表1に、各解析を行うために、着目したタグを示す。

表1 解釈タグリスト

①文書のサマリー	FONT,STRONG,B,H
②文書の内部構造	DIV, TABLE, OL, UL, HR, IMG
③文書の外部構造	A

これら着目するタグと解析機能の関係は、最終的な解析結果を左右するため、本来的にHTMLの仕様に含まれているものだけで

はなく、多くのWebページから実データを取得して、実際のタグの役割を検討する必要があると思われる。実際に、おそらくはWebデザイナーやWebページ作成ツールの問題もあるとは思われるが、タグの使用に関してサイト毎の傾向も見られたのは、興味深い発見であった。

また、本実験システムのユーザインタフェースのイメージを図2に示す。

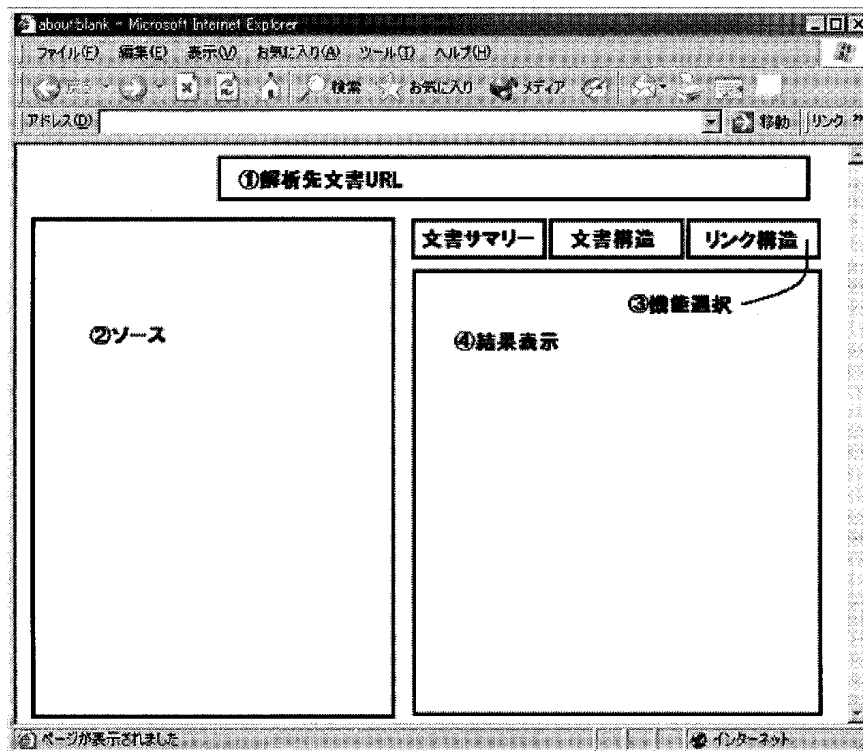


図2 ユーザインタフェースイメージ

- ①解析先文書URL
- ②文書ソース
- ③機能選択ボタン
- ④結果表示領域

①はユーザにより入力され、その結果得られた文書のHTMLソースが②に表示される。Webサーバよりエラーが返ってきた場合、エラーコード及びエラーの理由が表示される。②に対する解析処理を、③により選択する。最終的な解析結果が、④に表示される。

なお、本試作システムの開発は、ソフトウェア開発企業の新入社員教育の一環として行ったため、他に以下のような開発要求を挙げた。

- Java言語を用いて開発すること
- サーバ上で稼動し、複数クライアントからの要求に答えること
- クライアント側では、特定のクライアント環境を必要としないこと
- システムの拡張に耐えられること

クライアント側のエージェントソフトに依存しないシステムとするため、「特定のクライアント環境を必要としない」といった要件を挙げたため、実システムはアプレットを使わないサーブレットとして開発した。

4. 文書の解釈実験結果と考察

以上述べた仕様に基づき開発したシステムを使って、多くのサイトに対して解析実験を行った。システムの趣旨から、図や動画が多用されているものよりも、文書そのものが多くの文字情報から成るページの方が、かなり顕著な効果を上げた。

そういった特徴を持ったページの典型例として、官公庁の様々な公文書や官報などが電子情報として公開されているサイトがある。検討例として、文部科学省による「体系的な情報教育の実施に向けて（平成9年10月3日）」(<http://www.monbu.go.jp/singi/chosa/00000098/>)を取り上げた(図3)。当該ページは、文字数として25772字を含み、画像などは含んでいない。またページの先頭部分に、ページそのものの目次が示されているため、サマライズや文書構造の解析結果と比較し易いと思われる。

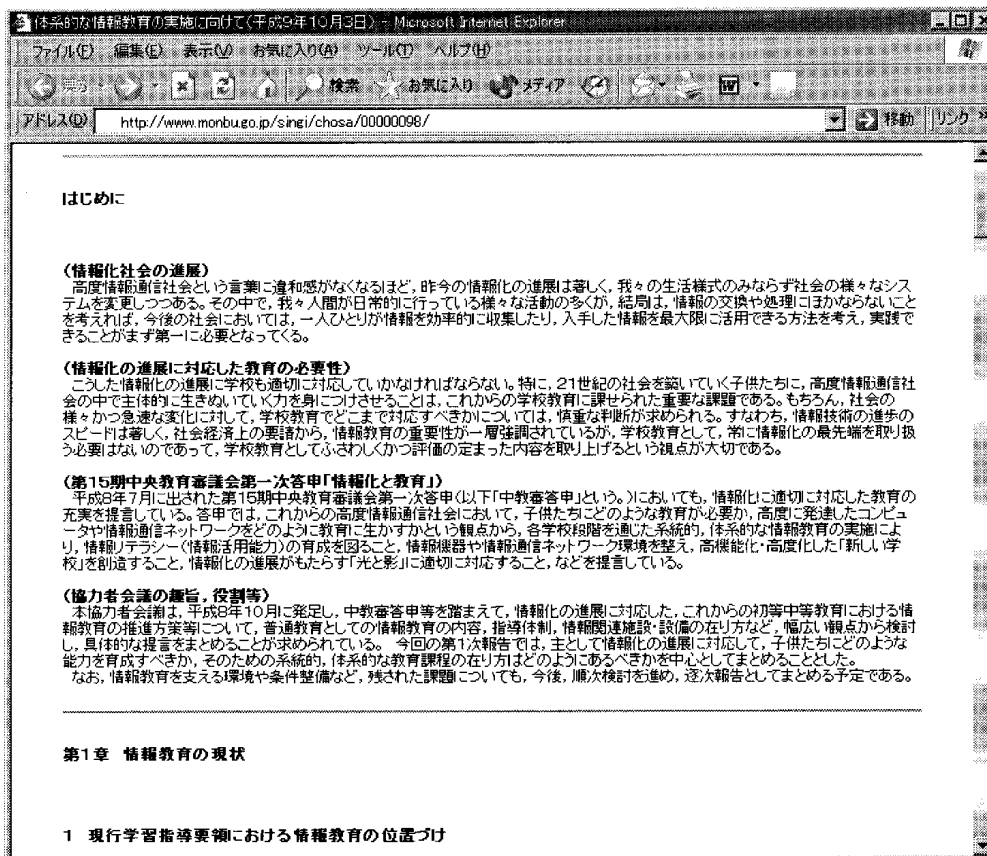


図3 解析実験対象サイト

本文書の解析結果を、資料A、B、Cに添付し、また画面出力イメージを図4、5に示す。表2には、サマリー処理に関して、当該ページの先頭に示されている目次項目と、本実験システムによるサマリー解析結果の比較を示す。これは目次項目毎の、システムの処理結果を示したものである。

単純に文字数としての比較はできないが、文書全体では25772字から1204字を抽出しており、要約率としては約4.6%となる。本文書では、強調「」が多用されているため、ほぼ目次項目の詳細と一致しているが、文書によっては字体「」や「B」なども抽出できると思われる。

さらに、各章毎の文字数とサマリー結果は以下の通りである。文書全体に渡って、ほぼ同程度の要約率を保っている。

- はじめに
文字数：1156、抽出数：69
- 第1章 情報教育の現状
文字数：5584、抽出数：265
- 第2章 これからの学校教育の在り方と情報教育の役割
文字数：10530、抽出数：428
- 第3章 次期学習指導要領の改訂に向けた提言
文字数：5087、抽出数：263
- 今後の検討課題等
文字数：1295、抽出数：50

なお章毎の解析結果の計と文書全体の解析結果計の数値に違いがあるが、これは文書タイトルや項目名そのものなどを算入しているためである。

ここに見る限り、予想以上に明確な解析結果が出力されたと思われる。出力結果を見ると、本来の意味でのサマリーとは言い難い部分があるが、文書全体の概要は十分把握することが出来る。特に興味深いのは、タグによって抽出した部分で用いられている部分に含まれている語彙が、文書上も重要なものであり、頻出語彙の上位を占めるという点である。前述のように、先行技術である自然言語処理を統合させるにおいては、辞書や語彙データベースなどの抽出のために、タグ情報を利用することができると思われる。

解析結果のより詳細な分析、評価のためには、さらに着目するタグとの相関をより深化させねばならないであろう。また本研究の前提と逆行するようではあるが、Web文書によっては、文書の装飾自体に文書の意味を反映させたデザインもあり、HTMLから装飾要素を分離するスタイルシートなども含めて解析に用いる必要があるように思われる。



図4 文書のサマライズ出カイメージ



図5 文書構造解析出カイメージ

元々タグ情報に着目したのは、Webの検索システムの仕様を検討している際に、殆どのサーチエンジンが、文書の意味情報そのものには立ち入っていないということが明らかになったことによる[1]。現在のサーチエンジンは、おおまかには、検索キーの頻度に着目するものと、文書のリンク構造に着目するものがあるが、どちらも本来的な意味で、文書の持つ意味情報を扱うものではない。

また[7]では、「従来の情報検索システムにおいては、システムと利用者に焦点が置かれ、検索の対象となる情報そのものにはそれほど注意が払われていなかった」と指摘されている。これは出版物を中心とした情報の検索における指摘であるが、Webにおいても状況は同じである。Webには、その発信される情報をチェックするようなシステムや組織が存在しないため、その内容の質を保っていくことは非常に困難であるという事情も存在する。

そのため、[7]で指摘されるように「サーチエンジンによる検索結果はそのサーチエンジンそのものの機能や利用者の検索戦略によってのみ影響を受けるのではなく、検索の対象となる情報そのものも、検索結果に少なからず影響を及ぼしている」と考えざるを得ないだろう。また[8]では、「情報発信をしたい、あるいは検索して欲しいキーワードを繰り返しを避けて多種類使用するなど、高得点で検索されるような工夫が必要である」と、情報発信をする側が発信の仕方を工夫することで検索結果の改善に貢献できるといったアイデアが明らかにされている。筆者も[1]で、同様の報告をした。「検索」処理を単なる情報処理ではなく、コミュニケーションと捉える考え方である。

こうした一連の指摘や先行研究に基づき、筆者らはタグ情報を用いた検索エンジンの仕様化と試作を行ってきたが、その機能モジュールの一つとして、意味解析をする本システムのアイデアが浮上してきた。

もしタグ情報を用いた文書の検索や意味情報解析が有効であると

するならば、より有効性の高いサーチエンジンの開発が可能である。また、情報の発信側に対して、タグ使用に関する一定の基準を与えることにもなり、Web情報の精度を上げる効果も期待できる。

さらに、大量のWebデータに意味を与えて、データベース化するという応用可能性も考えられる。筆者が[9]で指摘したように、Webからは「多くの情報を得ることができるように思えるが、実際には企業側が利用できるような整構造のものではない。そのため、企業にとって貴重な消費者からの生の声を大量に集めたデータベースが、収集、蓄積、再利用などの対象とは全くなっていない。」といった問題に対する、一つの解決策ともなる可能性をも持っていると思われる。

Webは人類の最大のデータベースである。社会的にはいろいろな問題点があるのかもしれないが、このことは否定できないだろう。であるとするならば、それを利用するための技術が、ネットワークそのものを高度化、効率化する技術に劣らず、必要であるのは間違いの無い事実である。

5. おわりに

開発に携わってくれた技術者達も本システムの仕様書で指摘しているが、現在までにタグに注目してWebページを扱うシステムは、今まで殆ど存在しなかった^[10]。そこにはいろいろな理由があるとは思われるが、おそらくタグが文書の意味構造を示すという、ごく当然の技術的前提が、技術者、研究者側にとっても蔑ろにされているというのが本当のところではないだろうか。

その意味では、本研究では当然のことを検証し、当たり前の結果を得られたに過ぎない。しかし、その当たり前のことが明確に検証できたということは、技術上の意義は決して小さくない。その技術の本来的な趣旨が、実際に歪むことなく使われているケースもあるということの証左でもあるがゆえである。個人的な感想

では、ここまで興味深い解析結果が得られるとは思ってはいなかった。予想がいい意味で裏切られたのはうれしい誤算であり、筆者らの技術観に誤りが無かったとも言えるであろう。

筆者のアイディアに基づいて、本システムの試作に携わってくれた、株式会社リバティシステムの若手技術者諸君、古賀・谷畑・土谷・中西の各氏に感謝します。君たち「春木組」が、新しい時代の技術者として活躍することを期待しています。

なお既存の学内ネットワークにセキュリティ上の危険や負荷などを与えないように、本システムを稼働させている実験サーバを学外になんとか公開しようとしたが、残念ながら学内稟議が否決されてしまったために、それを実現することができなかった。本研究のアイディアはいくつかのプロジェクトなどから関心を持ってもらったのではあったが、そうした事情から以降の展開が頓挫しているのは、研究者として残念である。実験結果からは、特に電子政府などの行政サービスの一環として有効性を持つのではと思われるのではあるが。

システムの管理とは、利用者に制約を与えることではなく、利用者に最大限の便益を与えることを意識すべきではなかろうか。少なくとも正当な技術的観点から、システムの運営・管理方針を立てるべきと考える。システムが、本当の意味で利用者のために機能すべきことを願い、本稿のまとめとする。

【参考文献】

- [1] 春木良且：「知的サーチエンジンの開発に向けて一分散環境における知識の構造とその再利用に関する一考察」，国際交流研究，フェリス女学院大学国際交流学部紀要，Vol.3
- [2] 古瀬幸広,広瀬克哉：「インターネットが変える世界」，岩波新書，1996
- [3] 「Webブラウザ新時代 互換性問題を検証する」，日経BYTE，日経BP社，2001. 6, pp108～115
- [4] "W3C Markup Validation Service", <http://validator.w3.org/docs/help.html>

- [5] <http://web.kyoto-inet.or.jp/people/cozy-p/mult.html> など
- [6] 標準化推進委員会・標準化動向調査部会：「ドキュメント・データ標準の課題WG 報告書」，社団法人日本情報システムユーザ会，1995
- [7] 古村哲夫：「WWWの情報検索システムの変遷ーサーチエンジンを中心にー」，<http://www.slis.keio.ac.jp/~ueda/sotsuron96/komura96.html>
- [8] 高木義和：「インターネットにおける情報検索」，情報管理，Vol.38, No.10, p.891-900, Jan. 1996
- [9] 春木良且，福田浩之：「ネット上の口コミをどうする」，第一回情報セキュリティシンポジウム発表資料，2004
- [10] 株式会社リバティシステム：「Web文書解析システム（HWDAS）機能仕様書」，2005

【資料】文書解析結果

章立て	目次項目	解析結果
	はじめに	
		(情報化社会の進展) (情報化の進展に対応した教育の必要性) (第15期中央教育審議会第一次答申「情報化と教育」) (協力者会議の趣旨，役割等)
	第1章 情報教育の現状	
	現行学習指導要領における情報教育の位置づけ	
		(「情報活用能力」) (「情報活用能力」の具体化) (現行の教育課程上の情報教育の扱い) (効果的学習指導のための情報手段の活用)
	情報教育の実施状況	
		(小学校の状況) (中学校の状況) (高等学校の状況) (特殊教育諸学校の状況)
	情報教育を進める上での課題	
		(情報教育の内容とその体系の明確化) (学習指導要領上の課題) (指導者側の課題) (学校における情報環境整備上の課題)
	情報教育の内容の体系化の視点	
	第2章 これからの学校教育の在り方と情報教育の役割	
	情報教育の目標	

		(情報教育の目標) (「生きる力」の育成と「情報活用の実践力」) (情報活用の基礎・基本となる「情報の科学的な理解」) (健全な社会建設のための「情報社会に参画する態度」) (3つの能力の関連性)
発達段階に応じたカリキュラム編成		
		(小学校の発達段階) (中学校の発達段階) (高等学校の発達段階)
情報教育の体系		
		(「情報活用の実践力」の育成と学習活動の範囲) (「情報の科学的な理解」の扱いと範囲) (「情報社会に参画する態度」の扱いと範囲) (情報手段を活用する上での留意点)
情報教育と教科の枠組み		
		(教科の枠組みの考え方) (クラス担任制と教科担任制) (情報に関する独立教科を設置することの効果等) (独立教科を設けた場合、それを必修にすることの効果等) (小学校段階での検討事項) (中学校段階での検討事項) (高等学校段階での検討事項) (特殊教育における検討事項)
第3章 次期学習指導要領の改訂に向けた提言		
各学校段階における情報教育の実施に関する基本的な考え方		
		(情報教育のための教科構成の在り方) (次期学習指導要領を検討する上での前提条件) (小学校段階) (中学校段階) (高等学校段階) (特殊教育諸学校) (各学校段階を通じて)
既存教科等における「情報活用の実践力」の育成		
		(既存の教科での「情報活用の実践力」の育成)
今後の情報教育の展開に向けて		
		(定期的な内容の見直しの必要性) (担当教員の養成及び研修) (将来の各教科の在り方の検討の柱の一つとしての情報教育)
今後の検討課題等		
		(指導の効果を高めるための情報手段の活用) (学校教育の質的改善) (学校の高機能化)

-
- 強調 : (指導者側の課題)
 - 強調 : (学校における情報環境整備上の課題)
 - 項目3 : 4 情報教育の内容の体系化の視点
 - 項目3 : 第2章 これからの学校教育の在り方と情報教育の役割
 - 項目3 : 1 情報教育の目標
 - 強調 : (情報教育の目標)
 - 強調 : (「生きる力」の育成と「情報活用の実践力」)
 - 強調 : (情報活用の基礎・基本となる「情報の科学的な理解」)
 - 強調 : (健全な社会建設のための「情報社会に参画する態度」)
 - 強調 : (3つの能力の関連性)
 - 項目3 : 2 発達段階に応じたカリキュラム編成
 - 強調 : (小学校の発達段階)
 - 強調 : (中学校の発達段階)
 - 強調 : (高等学校の発達段階)
 - 項目3 : 3 情報教育の体系
 - 強調 : (「情報活用の実践力」の育成と学習活動の範囲)
 - 強調 : (「情報の科学的な理解」の扱いと範囲)
 - 強調 : (「情報社会に参画する態度」の扱いと範囲)
 - 強調 : (情報手段を活用する上での留意点)
 - 項目3 : 4 情報教育と教科の枠組み
 - 強調 : (教科の枠組みの考え方)
 - 強調 : (クラス担任制と教科担任制)
 - 強調 : (情報に関する独立教科を設置することの効果等)
 - 強調 : (独立教科を設けた場合、それを必修にすることの効果等)
 - 強調 : (小学校段階での検討事項)
 - 強調 : (中学校段階での検討事項)
 - 強調 : (高等学校段階での検討事項)
 - 強調 : (特殊教育における検討事項)
 - 項目3 : 第3章 次期学習指導要領の改訂に向けた提言
 - 項目3 : 1 各学校段階における情報教育の実施に関する基本的な考え方
 - 強調 : (情報教育のための教科構成の在り方)
 - 強調 : (次期学習指導要領を検討する上での前提条件)
 - 強調 : (小学校段階)
 - 強調 : (中学校段階)
 - 強調 : (高等学校段階)
 - 強調 : (特殊教育諸学校)

横罫線

項目3 :はじめに

横罫線

項目3 :第1章 情報教育の現状

項目3 :1 現行学習指導要領における情報教育の位置づけ

項目3 :2 情報教育の実施状況

項目3 :3 情報教育を進める上での課題

項目3 :4 情報教育の内容の体系化の視点

横罫線

項目3 :第2章 これからの学校教育の在り方と情報教育の役割

項目3 :1 情報教育の目標

項目3 :2 発達段階に応じたカリキュラム編成

項目3 :3 情報教育の体系

項目3 :4 情報教育と教科の枠組み

横罫線

項目3 :第3章 次期学習指導要領の改訂に向けた提言

項目3 :1 各学校段階における情報教育の実施に関する基本的な考え方

項目3 :2 既存教科等における「情報活用の実践力」の育成

項目3 :3 今後の情報教育の展開に向けて

横罫線

項目3 :今後の検討課題等

横罫線

項目2 :要項&メンバー

桁揃え = right :平成8年10月4日

桁揃え = right :初等中等教育局長裁定

横罫線

横罫線

③リンク構造

文書タイトル :体系的な情報教育の実施に向けて (平成9年10月3日)

リンク :

⇒../index.htm#gaiyou

リンク :はじめに

⇒#01

-
- リンク :第 1 章 情報教育の現状
⇒#02
- リンク :第 2 章 これからの学校教育の在り方と情報教育の役割
⇒#03
- リンク :第 3 章 次期学習指導要領の改訂に向けた提言
⇒#04
- リンク :今後の課題等
⇒#05
- リンク :<H3>要項&メンバー</H3>
⇒#06
- リンク :ページの先頭へ
⇒#top