

ITC WebサイトのXMLデータベースを用いた リニューアル

■ITC本部 細川 達己

1. 従来のITC Webサイトの問題点

ITCのWebサイトに関しては、以前から多くの問題点が指摘されていた。これについてはITC内部からのみの意見ではなく、各地区ITCの旧Webサイトでも利用者対象のアンケートを実施したうえで、検討を行った。

1.1. 各地区ITC Webサイトのデザイン・操作性の不統一

実は前回（2006年）のリニューアルでもデザイン・操作性の統一は目標とされたが、統一する事自体に合意が取れていなかったことや、デザイン等に各地区からの意図があまり反映されていなかったことから、結果的に全サイトのデザイン・操作性を統一することはできなかった（図1）。



〈図1 旧各地区ITC Webサイトのデザイン〉

また、全体に共通するkeio.jp等の情報はITC本部のサイトに置かれたため、リンクをたどると全くデザインや操作性の異なったページに飛ぶようになっていた。

1.2. 目的とする情報の発見しづらさ

全体的に階層構造が深く検討されておらず、きちんとした階層メニューやいわゆるパンくずリスト（画面上部などに表示されるそのコンテンツまでの階層構造を含むリンク）なども用意されていない地区ITCサイトが多かった。さらに、検索窓やサイトマップなども用意されていない地区ITCサイトが多かったため、利用者が目的とする情報を発見することが困難だった。

1.3. コンテンツ共有の困難さ

各地区で共通なサービスの使用法など、サイト間で共有できそうなコンテンツが以前か

ら多くあったが、実際には共有されず、各地区でそれぞれさまざまなコンテンツを作成していた。

また、仮にカット&ペースト等の方法でそれらの内容を共有しても、元のサイトが更新された場合に、それに簡単に追従することが難しかった。一方これを、単なる他地区サイトへのリンクにした場合は、先述の「全く異なったデザイン・操作性のサイトへのリンク」の問題が発生してしまっていた。

1.4. ニュース作成画面の機能不足

前回のリニューアルのシステムでは、簡単なニュースであればHTMLを書かずに入力できる仕組みが提供されていたのだが、ニュース記事にしか利用できず、それ以外の一般の記事を書くことができなかった。

また、文書中でリンクを自由に張ったり、画像を埋め込んだりすることができず、結局大多数のページについては、各スタッフがそれぞれ自己流でHTMLを書く体制となってしまうていた。

1.5. 各スタッフがHTMLを記述することによる問題

各地区ITCサイトには、様々なスタッフが様々な時期に様々なスタイルで書いたHTMLファイルが大量に溜まっており、この先放置する時間が長くなればなるほど、そのようなファイルがさらに溜まっていくことが明らかだった。

また、HTMLの記述自体も、現在のWebサイトとしては問題のあるコンテンツが多数存在しており、文章内容とスタイルの記述が非常に混在した、再利用性の乏しいHTMLファイル群であった。

1.6. 認証付きページの扱いの問題

前回のリニューアル時点では認証付きページが想定されていなかったため、その後場当たりにkeio.jp認証を提供する別のサイトに一部のコンテンツが置かれることになったが、編集・管理が困難なものとなっていた。

1.7. 多国語化対応

多国語化に関しては、各地区で部分的に通常のコンテンツの一部としての対応を行っていたが、元々のサイトのテンプレートが多国語化対応しておらず、また、日本語コンテンツとの対応を取りながらメンテナンスしていくことが困難だった。

1.8. デザイン上の問題

画面幅を固定したいいわゆる「ソリッドレイアウト」である上に、画面幅の想定が狭く、また全体が極端に画面左に寄っていたため、利用者からの不満が多かった。

2. 現在の義塾全体のWebサイトの問題点

一方で、現状の義塾のトップページは、WebRelease2という商用のCMSで管理されており、この利用も検討してきた。しかし、以下の問題があるため、ITCとしては技術的・運用的な実験も兼ねて、新しい独自のシステムを構築することとした。

2.1. 解決できると思われる問題が限定的

スタイルの統一問題、デザインの問題、個々のスタッフでHTMLを書くことにまつわる問題はおそらく解決ができると思われるが、それ以外の問題に関してはあまり解決が期

待できない。

2.2. コンテンツ再利用性の問題

ページ内の構成という面では、ある程度意味的な部分もテンプレートに加えることができるが、文書の編集はWYSIWYGエディタ的な面が強く、文書中の個々の部分に意味を補足するようなマークアップはなされていない。また、他のシステムにコンテンツを移す際などに利用できるだろう、文脈的な意味を保った形のエクスポートも、標準ではサポートされていない。

3. 問題点の解決に向けて

以上のような問題点の解決に向けて、ITC Webのリニューアル作業を開始した。

3.1. 体制

前回のリニューアルの反省を踏まえ、全地区からメンバーを出すワーキンググループを作り、そこで議論しながら基本的な設計を行った。特に、システムの根幹に関わる部分はすべてITC内部で設計を行うこととした。

3.2. XMLの利用

コンテンツの再利用性を改善するため、HTML/XHTMLでコンテンツを持つのではなく、コンテンツはすべてXMLとして記述するとともに、XMLの知識なしでもある程度のWebページを作成できるようにする編集システムを用意することとした。

3.3. コンテンツ設計

全地区ITCサイトの基本的なコンテンツの構造を可能なかぎり統一した上で、これらのサイトを全てまとめて管理することとした。この複数のサイトの集合を本システムでは「ゾーン」と呼ぶ。ゾーン内ではサイト間でコンテンツの共有が可能（一つのコンテンツを複数のサイト内に表示できる）であり、サイト間での内容の差分も記述可能である。

文書、メディアファイルなどのコンテンツは全て「オブジェクト」として扱う。オブジェクトはオブジェクトIDという、ゾーン内で固有の文字列を持ち、サイト内リンクなど、他オブジェクトの参照はURLではなく、すべてオブジェクトIDを用いて行う。

さらに、コンテンツの階層構造を記述する特殊なリンク形式を用意して、自動的に階層構造をシステムが検索できるようにした。サイトマップや階層メニューはこの構造から自動生成される。

そして1つのオブジェクトに多言語のコンテンツを格納できるようにすることで、統一的に多言語コンテンツを提供する。

以上の機能を用い、サイト構成を部署視線ではなく、ユーザ視線でまとめ直すこととした。ユーザの多くは所属地区でのIT活用に必要な情報を求めているのであり、「各地区ITC」や「ITC本部」のサイトを求めているわけではない。

そこで、各地区向けのサイトで、その地区内で必要なITCサービスに関連する情報が、サイト内ですべて完結する形で表示されるようにした（例えば、keio.jpのように全地区で共通のサービスに関するページも、コンテンツ共有機能を用いることでそれぞれの各地区向けサイト内に取り込んだ）。

そして、必要なくなった「ITC本部」のサイトやkeio.jpを解説するサイト等は廃止し、

一方新たにITC全体の外部向けサイトを用意することとした。

3.4. システム設計

コンテンツをXML形式にしたことで、XMLデータベースに全コンテンツデータを格納し、コンテンツに対する様々な横断的検索が可能となる。そして、そのXMLデータを表示プログラムによって動的にXHTML、もしくは適切な形式に変換して表示する。

これらのXMLコンテンツの検索・管理などを行うことができるインターフェースとして、簡単なCMS機能を提供した。スタッフが仮にXMLを記述できなくても、通常の記事を書くには十分な程度の自由度・表現力を持つ編集機能を、このCMSは提供する。

アクセス制限に関しては、keio.jpによる認証ページ（ただし塾内からは認証なしで閲覧可能）を系統的に簡単に提供できるようにした。

3.5. デザイン・ユーザビリティの改善

全地区ITCサイトのデザイン・ユーザビリティを統一した上で、以下の基本方針で改善を行うこととした。

3.5.1. 統一したナビゲーションの提供

基本的な階層を簡単に確認できるメインメニューを提供し、また階層に対応するいわゆるパンくずリストを表示する。さらに検索窓も提供する。

3.5.2. リキッドレイアウト

固定的な画面幅を想定せずに、画面の幅をあるだけ有効活用する、いわゆる「リキッドレイアウト」のデザインとした（図2）。広い画面では広さを有効活用可能であり、一方狭い画面でも横スクロールを必要としない利点がある。ただしその反面、デザインには制約事項が多く、難易度は高い。



狭い画面

広い画面

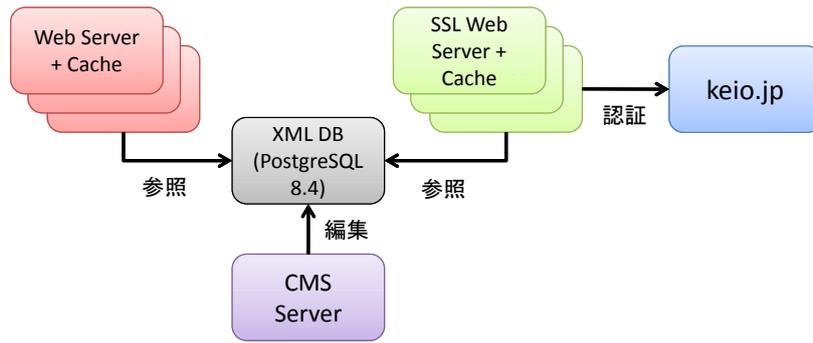
〈図2 新ITC Webサイトのリキッドレイアウト〉

4. 新ITC Webシステム

以上の目的を達成するため、次のような新ITC Webシステムを構築した。

4.1. システム全体像

XMLデータベースサーバと、そこに接続する複数のWebサーバ、そしてCMS用のサーバで構成される（図3）。XMLデータベースとCMS用サーバは同一マシンで兼ねることが可能である。



〈図3 ITC Webサイトのシステム構成〉

Webサーバは通常コンテンツ用サーバとkeio.jp認証コンテンツ用サーバの二種類があり、どちらも複数化して負荷分散することが可能である。性能向上とデータベースサーバの負荷軽減のため、各公開用Webサーバにコンテンツのキャッシュを持つ。

全サーバは仮想化システムに置かれており、停電時の拠点間移動や、負荷増大時にWebサーバを追加することなどを容易に行うことができる。

4.2. XML設計

本システムのXML仕様は、現在のところ、XHTMLから受け継いだ基本的要素21個（
<lockquote>
<cite><dd><div><dl><dt><kbd><noscript><p><pre><q><script><sub><sup>）と、本システム用に独自に追加された要素37個で構成されており、登録時に文法チェックが出来るようにXML Schemaも定義されている。そのため、登録されているXMLデータは、すべて規定の文法に従っていることが保証される上、HTMLのような曖昧さはない。

XHTML で表示されるWebページや画像ファイルなどのバイナリファイルは、先述のように「オブジェクト」として定義される。オブジェクトはそれぞれXMLデータ1つに対応している。

共有するサイト別の差分記述などを実現するために、特定のサイトのみでXMLデータの一部を有効にするような文法を用意した（図4）。

```

<p site="sfc">SFCでのみ表示される段落</p>
<def key="campus" site="mita">三田</def>
<def key="campus" site="hc">日吉</def>
<p><ref key="campus" />キャンパス</p>

```

任意のタグに付けられるsite属性によって、サイトごとの場合分けが実現できる。
def要素とref要素で、キーワードの定義・置換も可能。

〈図4 共有コンテンツにおけるサイトごとの場合分け〉

これを高速に検索するために、オリジナルのXMLデータに対して、サイトごとの前処理を行い、各サイト用の個別のXMLデータに変換してからデータベースに格納している。この機能は、後述の協調支援にも利用される。

1つのXMLデータに多言語のコンテンツを同居させることが出来る。これによって、リンク先のコンテンツがどのような言語集合で記述されているかも、簡単に検索が可能と

なる。

4.3. XMLコンテンツ仕様の拡張

文法を厳密にチェックし、限られた記述法しか許さないXML文書は、文書の再利用性を増すが、自由な形式による表現ができないと考えられる一面がある。本システムにおいては、スタイルシートをstyle属性でタグに指定することで、ある程度この問題を解決することも可能ではある。しかし、「このような表現がしたい」という要望があった場合は、検討の上、必要であれば新たにそのための要素をXML仕様に追加し、表示システムにも手を加えることによる対応を、本システムでは推奨している。

このように事後に追加された要素には、湘南藤沢ITCからの要望で作られた、定型文書ニュース用のタグなどがある。これは、「記」や日付、広報対象などの文書パーツを定型文書としてレイアウトするものであり、トップページに自動的に表示されるサマリーなどでは、本質的な広報内容に関係のない部分は表示から除外される。

このような、需要に応じた要素を提供することによって、文章の再利用性を高め、なおかつ簡単に定形文書ニュースを記述できるようになっている。

4.4. XMLデータベース

本システムでは、PostgreSQL 8.4のXML型サポートを用いて、XMLデータベースを構築している。ただし、当初はバイナリファイルのデータもbase64などでエンコードしてXMLとして格納することを検討していたが、これによって各XMLのサイズがメガバイト以上の単位となると、インデックスを用いた場合でもデータベースの性能が致命的に悪化することが判明した。そのため、バイナリデータは、メタデータをXMLとして格納する一方、バイナリデータそのものはPostgreSQLのBYTEA型（バイナリ型）で保存している。また、textsearch_sennaを用いた、XMLの高速な全文検索も可能である。

4.5. CMSシステム

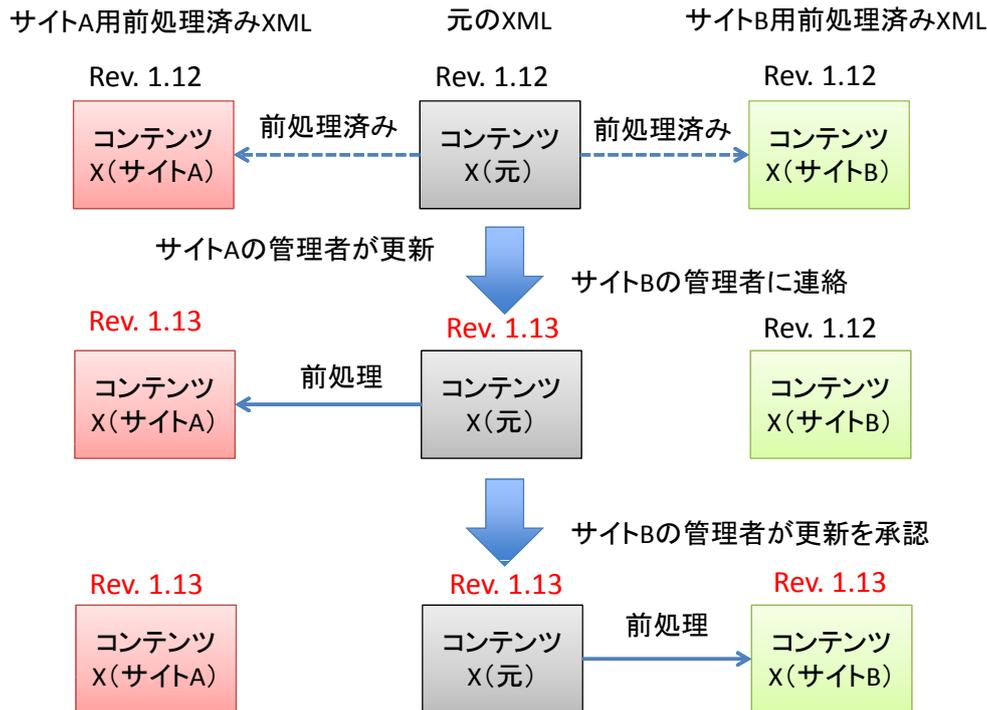
オブジェクトの検索、新規登録、更新、削除（実際には消さずに無効フラグを立てるだけ）が可能な、簡易CMSシステムを提供している。比較的単純なコンテンツであれば、XMLを書くことなくWebページ（に対応するXML）を作成できる編集機能も提供される。各オブジェクトに関しては、更新履歴を管理できる。この履歴管理に用いられるリビジョン番号は、後述の協調支援にも利用される。

4.6. 協調支援

本システムでは複数サイトで1つのオブジェクトを共用できるように設計したため、望まない変更が他サイト管理者によってなされる可能性が出てくる。

これを本システムでは、先述の「サイトごとに前処理されたXML」と「履歴管理」機能を用いて解決している。つまり、個々のサイトで独立した前処理済みのXMLから表示用のXHTMLがそれぞれ作成されることを利用して、サイトごとに異なったりビジョンのXMLを基にした表示を可能としている。

この仕組みを、図を用いて解説する（図5）。



〈図5 共有コンテンツ更新時の処理〉

仮にコンテンツX（リビジョン番号1.12）に対してサイトAの管理者が更新（リビジョン番号1.13に更新）を行った場合、サイトAを表示する際に使用されるコンテンツXのリビジョンは自動的に1.13となるが、サイトBはリビジョン1.12をそのまま使い続ける。

サイトBの管理者に対しては、サイトAによる更新内容がメールで知らされる。問題がない場合は承認の処理をCMSシステムから行うことで、サイトBもリビジョン1.13を用いることになる。

サイトBに異議があった場合は、更新を拒絶するのではなく、新しい更新をサイトBによって起こすこととした。その場合、今度はAが承認を行う側となる。

なお、当然このプロセスは続く可能性があるが、サイト間の差分記述が可能であることと、ゾーン内は関係の深いサイトのみが存在すると仮定しているため、そこまで「こじれる」ことは想定していない。

また、それほど重要な変更でない場合は、この協調支援機能を利用せずに承認不要で更新することもできる。各サイトは、「未承認の更新が存在しなければ最新のリビジョン、もしくは最も古い未承認の更新の1つ前のリビジョン」を表示するようになっているため、この承認不要の更新が、それより古い承認依頼を無効化することはない。

4.7. 表示システム

表示システムは、Apache HTTPサーバ上で動作する、FastCGIを用いたPerlのスクリプトで実現されている。複雑なXMLファイルをXHTMLに変換する場合や、バイナリファイルをデータベースから取り出して画像などを表示する場合も、比較的高速な表示が可能となっている。

また、Webサーバ側にキャッシュ機能を持っており、一度表示したデータは原則として静的コンテンツと同等の性能で表示することが可能である。これはApacheのmod_rewrite機能を用いて「キャッシュファイルが存在すればキャッシュを表示、そうでない場合はFastCGIスクリプトを呼び出して表示」という処理の流れを実現している。

特殊な機能として、Microsoft Excelのファイル（バイナリファイルのオブジェクトとして扱われる）を直接Webページに埋め込みすることが出来るようになっている（表示システムがExcelファイルからXHTMLへの変換、埋め込みを行う）。罫線の有無、フォントサイズ、フォントスタイル、背景色などの再現も可能である。これによって、多くのCMSシステムで問題となっている、HTML表の作成が非常に簡単なものとなる。

4.8. キャッシュ管理システム

先述のキャッシュは、オブジェクトが更新された場合は、そのオブジェクト自身と、そのオブジェクトの状態変化が影響を及ぼす他のオブジェクトに対応したキャッシュエントリを削除する必要がある。

これを行うために、定期的にデータベース中のオブジェクトの更新状況をチェックし、オブジェクトの依存関係に基づき適切なキャッシュエントリを消去するシステムが各公開Webサーバで動作している。

4.9. デザイン

先述のように、デザインはリキッドレイアウトを基本に行った。このレイアウトファイルから作成したテンプレートファイルの中に、表示システムはXMLからXHTMLに変換したコンテンツを埋め込み、ブラウザに表示するようになっている。

5. コンテンツの移行

5.1. コンテンツの調査・再配置

今回のリニューアルに際しては、各地区ITC及びITC本部のWebサイトのコンテンツを調査し、それらを仕分けた上で、できるだけ各地区で同じような階層構造とトップメニューとなるような形で再構成した。

5.2. 多国語化

既に英語化されていたコンテンツを、今回のフレームワークの中で多言語オブジェクトとして再構成し、また新たに、必要性が高いと思われる多くのページの英語翻訳を行った。サイト間でコンテンツを共有していると、これらの翻訳の成果も共有することが出来る。

5.3. XMLへの変換

コンテンツのXMLへの変換は、一旦Perlスクリプトを用いてラフな変換を行い、それを手作業でチェックしてXMLの文法の修正やリンク構造の調整などを行った。そしてこれらを、バッチプログラムで一気に新システムのXMLデータベースに格納した。

6. リニューアルにあたる内部開発と外部委託

今回のWebリニューアルに関しては、内部開発と外部委託の切り分けは次のようになっている。

- ITC内部開発：サーバ構築、XML設計、XML Schema作成、DB設計、CMS用のDBアクセスAPI実装、表示システム実装、キャッシュ管理システム実装、コンテンツ調

査・再配置、コンテンツXML変換（第一段階）

- 有限会社ハビタス：テンプレートデザイン（全7サイト）、CMSシステム実装、コンテンツXML変換（第二段階）
- 株式会社日立テクニカルコミュニケーションズ：英語翻訳

7. 本番公開

リニューアル公開は、東日本大震災等の影響で2011年度にずれ込んだが、2011年7月14日に無事、完了した。

8. 今後の展開

システム設計上は、複数のゾーンを作成することが可能であったが、現在のITC Webサイトのシステムではそこまでは不可能である。また、CMSからのテンプレートの編集など、CMSシステムに関する要望も、実際に利用を始めると出てきている。

これらの機能を実現することで、ITC以外の他部署も、このシステムによるWebサイトの構築が可能となる。また、ITC内でのサイト管理の利便性も向上することが期待できるだろう。