

インターネットストリームメディアの 現在と将来

上村 圭介

1. はじめに インターネットとメディア技術
2. 動画・音声のためのメディア技術
3. インターネット上のストリームメディア
4. メディア技術の行方
5. おわりに代えて マークアップ言語によるメディアの統合

要旨

インターネットの次のアプリケーションは、動画・音声などのストリームメディアだと言われる。インターネット上のストリームメディアをめぐるのは、リアルネットワークス、アップル、マイクロソフトなどが、独自技術が無償で提供して標準化の競争を行なっているように見える。しかし、本稿では、インターネットストリームメディアという領域が、インターネットの制約を受けて、独自に発展してきたものであり、今後この制約がなくなれば、インターネットの制約を当面解消するために導入された経緯をもつ、これらの技術の意義は相対的に低下すると考える。

また本稿では、ストリームメディアの技術的独自性が低下する中で、各々のメディアを統合するアーキテクチャである、W3Cのマークアップ言語"SMIL"が、どのような役割を果たしうるのかを示す。

1. はじめに インターネットとメディア技術

コンピュータ上では、動画、音声、静止画、文字などは、全て符号化され、記憶、転送される。その際、それぞれのメディアの特性に着目し、もっとも効率的な手法が用いられる。これらの符号化技術には、適用領域ごとに長所・短所があり、一概にどの符号化が優れているとは言えず、ユーザがどの符号化技術を採用してよいのか決めにくいこともある。インターネット上での音声配信、画像配信のための技術は、一時期は非常に多様であり、ユーザは混乱してきた。しかし、ようやく最近になって、インターネットでのメディアのための符号化技術は整理されてきた。本稿では、インターネット上のメディア配信のための技術という観点から、これらの符号化技術の動向について概観する。

2. 動画・音声のためのメディア技術

動画・音声などのメディアは単純に符号化すると、データ量が膨大なものになる。そのような膨大なデータのままで、記憶装置の容量や通信回線の速度をこえてしまうことが多い。そのため、これらのメディアでは、従来から「圧縮」と呼ばれる手法によって記憶時や転送時のデータ量を縮小することが行なわれてきた。メディアの圧縮には、メディアそのものの特性、メディアが人間の感覚に働きかける際の特性、コンピュータ上での利用形態などの諸条件を考慮して、様々な符号化方式が開発され、利用されてきた。いずれの場合も主な動機は符号化を通して、データ量を縮小することである。

例として動画に必要とされるデータ量を示そう。表1は、VHSビデオ相当の画質のデータを記憶する場合に必要なデータ（無圧縮データ）と圧縮時のデータの概算量である。

時間	無圧縮のデータ量	圧縮時のデータ量
1秒	7.55 MB	0.19 MB
1分	452.73 MB	11.25 MB
1時間	27,163.97 MB	675.00 MB

表1 動画データ量の比較

表1に示した通り、音声や動画メディアのデータ量は非常に大きく、そのまま無圧縮で記憶または転送することは現実的ではない。無圧縮では、2時間の映画でも50GB以上のデータ量になってしまう。これでは、とてもデータの蓄積や転送に向かないということが理解できるだろう。「圧縮時のデータ量」は、同じ動画をMPEG-1で圧縮した場合のデータ量に相当するが、ほぼ40分の1になる。これであれば、1時間の動画でも約675MBのデータ量となり、CD-ROMに蓄積するこ

とが可能となる。

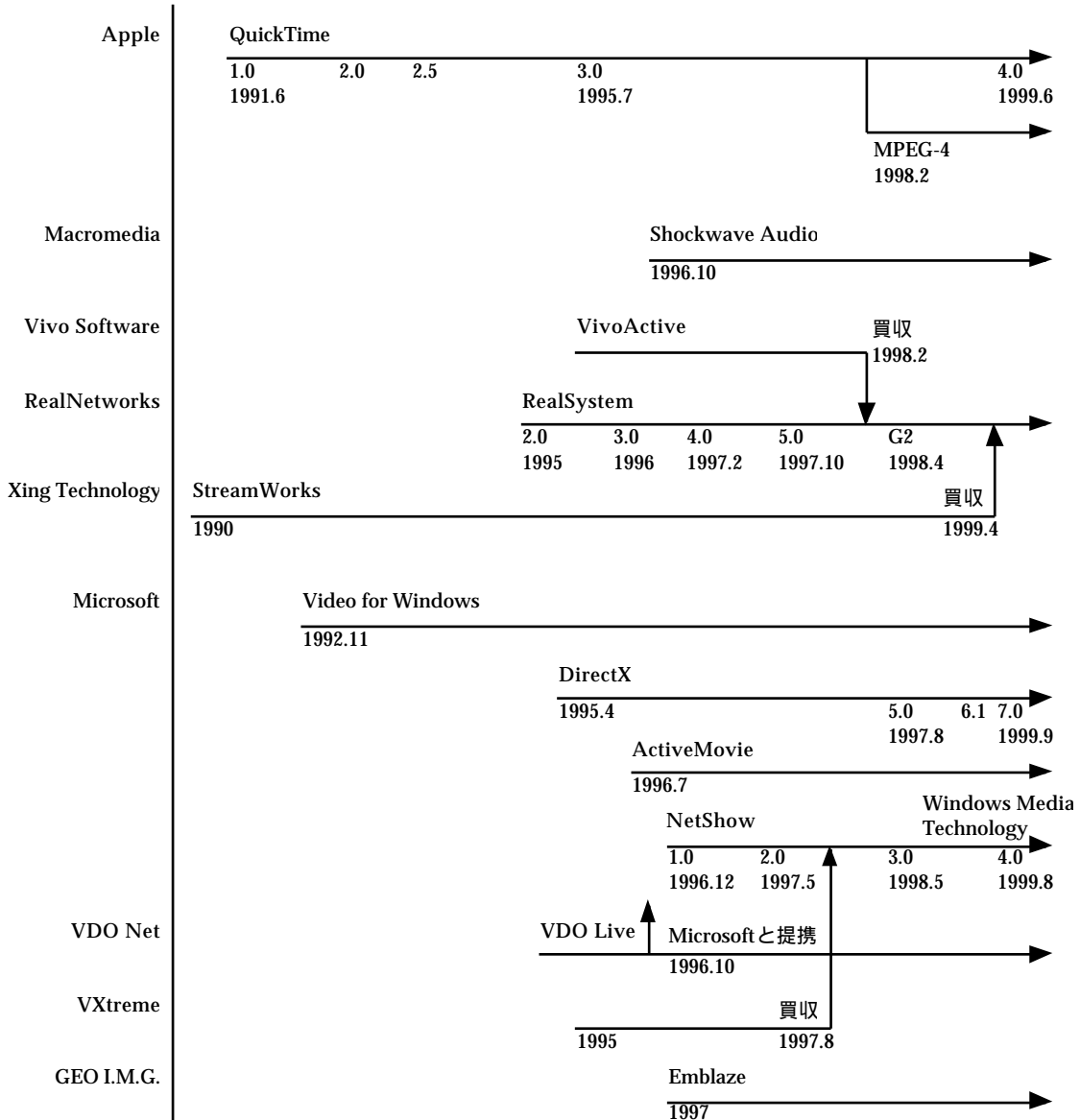
表1で示した値は、CD-ROMへの蓄積を適用領域としたMPEG-1の場合の数字であるが、他の適用領域（例えばテレビ電話）では、また異なった圧縮方式が開発され、採用されている。これは、適用領域ごとに必要な圧縮率などの条件が異なるためであるが、そのために音声、動画のための圧縮方式は、非常に多様である。

3. インターネット上のストリームメディア

インターネットの商用利用が始まり、インターネットを次のメディアとして利用する動きが世界中で始まると、インターネット上でストリームメディア¹を流す試みがなされるようになった。その過程で様々なベンチャー企業が立ち上がり、様々な特徴をもつ独自のストリーム送信システムを提供し始めた。その中には現在も主流技術として残っているものもあれば、競合他社に吸収されたものもあり、非常に流動性が高い。ここではまず、インターネットをターゲットとした音声・動画ストリーミングのための主要な技術の動向を図1にまとめる。

なお、ネットワークを介してビデオストリームを送出するシステムという点では、イントラネットを対象にしたものが多数市場に出回っている。しかし、本稿ではパブリックなコンテンツこそがインターネットのアプリケーションを形成するという立場をとるため、イントラネットをターゲットにした動画配信システムについては検討の対象外としておく。

¹ 「ストリームメディア」とは、時間軸に従って再生される音声や動画などのメディアのことをいう。ストリームメディアは内部に時間をもつため、クライアントソフトウェアがそのメディアを再生する場合には、メディアの全体が予め送信されている必要はなく、ある瞬間で再生するのに必要なだけのデータが、その瞬間に送信されていればよい。このようなデータの送信の方法は、特に「ストリーミング」と呼ばれ、ファイルの全体を予め送信する方法と区別される。



メディア技術の動向

これまでのインターネットストリームメディア技術は、メディアを送信するサーバ、受信するクライアント、サーバとクライアントの間でやりとりされるデータの形式と転送の際のプロトコルが、一つの「システム」のパッケージとして登場してきたものが多い。HTMLがそうであるように、データ形式さえ共通であれば、通信の両端（サーバとクライアント）には任意のソフトウェアを採用しても構わないはずであるが、実際にはそうになっていない。データ形式や通信プロトコルの基本的な部分は共通していても、細かな実装では異なり、任意のソフトウェアを採用できる水準の相互運用性は達成されていなかった。

このように関連するソフトウェアとデータ形式がひと括りに開発されてきたため、技術やソフトウェアの名称を明確に区別することが難しい場合がある。図1に示した技術も、ソフトウェア（パッケージ）の名称（StreamWorks、Emblazeな

ど)、ファイル形式やメディア記述形式の名称 (Video for Windows など)、メディア処理用APIの名称 (DirectX など)、またはこれらすべてをひと括りにしたメディア処理環境としての呼称 (QuickTime、Windows Media Technology など) などが混在している。また、リリースの時期によって、ある呼称で総括される技術の範囲が異なることもある。これらの整理は、いずれ機会を改めて行ないたい。

次節以降では、図1に示した技術のうち、次の三つを中心に、それぞれの技術の特徴をまとめる。

- RealSystem (RealNetworks)
- QuickTime (Apple Computer)
- Windows Media Technology (Microsoft)

3.1 RealSystem

RealSystemの特徴は、RealNetworks (1997年10月にProgressive Networksから社名変更) によって開発されているインターネットメディアの草分けとも言えるシステムである²。現在もっとも多くのシェアを有するシステムであると言える。RealSystemは、28.8Kbps程度の速度のモデムによる低速のネットワーク接続でも再生可能な高圧縮フォーマットと転送制御を提供してきたことである。1995年にRealAudio 2.0、1996年にRealAudio 3.0をリリースし、1997年2月には動画送信に対応したRealVideo 4.0をリリースした。1997年10月にリリースしたバージョンである5.0からは、製品名として「RealSystem」を採用し、現在にいたっている。

RealNetworksは、再生クライアントであるRealPlayer (バージョン3.0までは、RealAudio Player) 同時接続クライアント数を25件に限定したライセンス限定版のサーバを無償で公開し³、普及促進を狙う一方で、RealPlayerで視聴することのできるコンテンツの充実も積極的に行なっている。既存の映画や音楽業界などとの提携を通じて、音楽サイトなども積極的に構築している。RealNetworksのマルチメディアポータルサイトは「RealGuide」として知られている。

また、1998年4月にリリースされたRealPlayer G2では、W3C勧告となったマルチメディア技術言語「Synchronized Multimedia Integration Language (SMIL) 1.0」をいち早く実装している。

3.2 QuickTime

QuickTimeは、インターネットへの適用以前から開発されてきたApple Computerのマルチメディア技術である⁴。最初のバージョンであるQuickTime 1.0は、1991年6月に発表された。これは、MacintoshのオペレーティングシステムSystem 7.0上のアプリケーションで、動画や音声を取り込むことを可能にし、「マルチメディア文書」を作成するための基礎技術である。同時に、開発当初からプレゼンテーション中のメディアを時間軸上に同期させて再生するための同期制御機能ももっ

² <http://www.real.com/>

³ RealNetworksが同時接続可能なクライアントに制限のあるサーバを無償で提供したのは、RealServer 5.0からである。それ以前のサーバは、すべて有償であった。

⁴ <http://www.apple.com/quicktime/>

ていた。

QuickTimeは、当初はインターネットに特化したメディアではなく、コンピュータ上で動画・音声を一般的に扱えるようにするための基礎技術であったが、インターネット使用の拡大に伴い、インターネット上のメディア形式としての特徴を備えるようになった。1995年7月にリリースされたQuickTime 3.0からは、Apple ComputerのMacintoshシリーズで利用できるだけでなく、MicrosoftWindowsでもこの技術が利用できるようになった。また、QuickTime 3.0のマルチトラック⁵のためのファイル形式は、ISOの画像圧縮規格であるMPEG-4にも採用された。

本稿執筆時の最新バージョンであるQuickTime 4.0は、インターネットでのストリーミングにも対応している。また、MacOS X Server上で動作するストリーミングサーバのソースコードがオープンソース化されている点も注目される。MacOS X Serverは、BSD系オペレーティングシステムの派生形であるため、Linuxなど、他のUNIX系オペレーティングシステムで利用される可能性もある。

QuickTimeは、音声、動画とも多様なフォーマットに対応しており、元となるメディアの形式の区別を意識せずに編集や制作の作業できるため、コンテンツ作成の場で重宝されることが多い技術である。

3.3 Windows Media Technology

Windows Media Technologyは、Microsoft Windowsファミリに実装されている一連のメディア技術である⁶。Microsoftは、Windows NT Serverに統合可能なストリームメディアサーバとして、NetShow 1.0を1996年10月にリリースした。1997年5月には、NetShow 2.0、1998年5月にはNetShow 3.0がそれぞれリリースされた。マイクロソフトは、Video for Windows（ファイル形式）、DirectX（マルチメディア処理API）、ASF（ストリームメディア形式）などを個別に発表しているため、一連のメディア処理技術が構成する全体像を把握しにくい。これらのいわば集大成が、1999年4月にリリースされた、Windows Media Technology 4.0と言えるだろう（そのバージョンも、NetShow 3.0に続く、4.0が与えられていることからみても、明らかである）。

Windows Media Technology 4.0は、現在のところWindows NTでのみ利用可能であるが、ソフトウェアを無償で入手することが可能であり、Windowsシリーズのオペレーティングシステムにおける圧倒的なシェアを背景にユーザ数を拡大している。この技術は、比較的后発であるが、「インターネットラジオ」のような音声配信のアプリケーションとして積極的に普及促進活動を進めており、音楽サイトや放送局サイトのように一般的なインターネットユーザが日常的に接するサイトでは、RealMediaと同じ程度に普及しているのを見ることができる。

一部には、Microsoftはもはや、Internet Explorerの新バージョン開発に消極的だという観測もある。これは、多額の予算を投じてInternet Explorerを開発してきた最大の理由ともいえるべきネットスケープ社との競争が事実上存在しなくなったことからの推測であるが、だとすると、Windows技術に収斂するというシナリオと

⁵ 一つのファイルの中で、音声、アニメーション、動画などを別々の「トラック」として格納する方式。

⁶ <http://www.microsoft.com/windows/windowsmedia>

は別のシナリオが考えられるのかもしれない。

3.4 その他の技術

インターネットが普及する契機となった1995年の時点では、インターネットストリームメディアの製品の数も多かったが、図1で示した通り、その多くは競合他社と提携し、または競合他社に買収され、現在ではこの分野のソフトウェアのバラエティはかなり限られてきた。1990年からMPEG-1を転送するためのソフトウェアを開発・販売していたXing Technologyは、1999年4月にRealNetworksに買収された。また、VivoActiveを提供していたVivo Softwareも、同じく1998年2月にRealNetworksに買収されている。

やや異色とも言えるのが、Emblazeである。これは、イスラエルに本拠をおくGEO Interactive MediaGroupが開発したシステムの名称で、再生用クライアントアプリケーションをJavaアプレットとして配付するために、固有のクライアントソフトウェアをユーザが事前にインストールしておく必要がない。通常のインターネットストリームメディアでは、再生を開始する前に、ユーザが再生クライアントをインストールしなければならない。これは、サービスの提供側がユーザに対してコントロールを有していることを意味するが、Emblazeでは、Javaによりプラットフォーム非依存のアプレットとしてクライアントソフトウェアを配付することで、ユーザへのコントロールを最小限にすることに成功している。しかし、Javaを利用しているため、起動に時間がかかる、動作速度が遅く、高いパフォーマンスのPCを必要とするなど、Javaを利用したアプリケーションに共通した欠点をもつ。

4. メディア技術の行方

初期には、各ベンダーのソフトウェアは自社のフォーマットだけをサポートし、ユーザの困り込みを指向していたが、現行のメディアサーバやプレーヤは、複数のフォーマットに対応している。例えば、Microsoft Media Playerは、MicrosoftのビデオフォーマットであるAVIだけでなく、QuickTimeやRealMediaのファイルも再生することができる。RealPlayer G2は、RealMediaだけでなく、AVI、MPEG、QuickTimeなどのフォーマットが再生可能である。QuickTimeプレーヤは、QuickTimeムービーだけでなく、MPEGムービーの再生が可能である。また、VDO Playerでは、テレビ電話を実現するために、H.263⁷などをサポートしている。

今後すべてのメディア形式が統一され、サーバ、再生クライアントが単一の形式に従えば（究極的には単一のベンダーによるソフトウェアが使用されることになれば）、ユーザは個別のソフトウェアを用意する必要がなくなり、ユーザにとっての利便性は向上する。しかし、ユーザの利便性を高めるには、逆に、それぞれのサーバ、ブラウザが多様な形式をサポートするという方向もある。各社が互いのフォーマットをサポートするようになったのは、この現れだというのが筆者の

⁷ ITU-TSS (International Telecommunication Union, Telecommunication Standardisation Sector) が制定したテレビ電話のための画像圧縮規格。

解釈である。もしこの解釈が正しければ、インターネットの映像アプリケーションでは、既に広く採用されている複数の形式が存在するため、これらの形式が、特定のベンダーの一つの形式に収斂する動きは、一時的には弱まるであろう。

このような複数フォーマットの混在と、クライアント側での対応能力の向上は、どのような意味をもつだろうか。これまでのインターネットのメディア技術は、インターネットでの音声、動画という「ニッチ市場」を埋めるために登場してきた。インターネットで映像を流す場合には、非常に低い回線速度や、通信品質を想定しなければならないために、圧縮率をさらに高め、ネットワークの遅延などに対応できる堅牢性が求められた。MPEG-1など、国際規格などで制定されてきた形式は、少なくともインターネットの商用利用が始まった1995年から現在までのインターネットの上で、そのまま採用することが難しかったと言える。のちに、MPEG-4のような比較的低速のネットワーク上で使用することを想定した規格が制定されることになって、**「インターネット時間」と呼ばれるほど急速に発展するインターネットの要求には合致しなかった。**つまり、それまでに開発されていた音声・動画フォーマットが想定していた回線品質は、当時のインターネットの回線状況と比べればきわめて高いものであったといえる。そのために、インターネットでストリーミングを行なうためには、著しく低い回線状況でも利用可能なフォーマットを開発することが必要だったのである。

むしろストリーミングメディアの領域では、WWWのための比較的単純な実装とは異なり、高度な技術要素を必要とするため、これまでは開発者が個々にデータ形式を開発してきた。

しかし、回線速度の問題が解消されれば、独自の高圧縮技術を開発する必要もなくなり、圧縮のための独自フォーマットをめぐる競争はあまり意味をもたなくなるだろう。繰り返しになるが、これは、これまでのインターネットストリームメディア形式の意義の大部分は、回線速度のニッチを埋めることにすぎなかったのである。

既にオーディオメディアでは、「MP3」と通称される圧縮方式（MPEG-1 Audio Layer 3）が急速に普及している。MP3は、それほど高い圧縮率を実現するわけではないが、その代わりに高い音質を提供する。いくら質のよい音楽を提供できたとしても、圧縮率が回線速度に不釣り合いなほど低く、高速なネットワーク接続を必要とするのであれば、一般には普及しなかつただろう。MP3の成功は、極限まできりつめた圧縮をしなくても、ネットワーク上でのダウンロードやストリーミングが可能であるほど回線速度に余裕が出てきたことの現れとして理解することができる。

究極的には、圧縮をする必要さえなくなることも想定できる。本稿がとりあげているインターネットストリームとは利用場面が異なるが、ビデオ編集やオーディオ編集の場では、音声圧縮の問題は解消されている。音声は、かつてコンピュータ上で処理する場合でも圧縮をしていたが、現在は外部記憶装置が廉価に入手できるようになったことなどで、コンピュータ上では無圧縮のまま処理されることが多い。例えば、デスクトップ編集ソフトウェアでも、ビデオは圧縮して操作しているが、オーディオは無圧縮である。このことは、システムの制約がなくなれば、圧縮技術のように、システムの制約を回避するための技術の意義が弱まるということを示唆していないだろうか。

例えば、1996年10月22日の『Interactive Age』の記事では、Vivo Softwareの社長兼CEOであるStaffan Ericsson（当時）の言葉を次のように紹介している。⁸

第一世代のメディア配信では、ユーザはたった30秒のAVIやQuickTimeファイルを30分かけてダウンロードしなければならなかった。第二世代では、Progressive NetworksのRealAudioのような独自技術によるストリーミング配信のための製品が出回るようになった。第三世代では、標準化されたサーバによってストリーミングメディアを配信することになるだろう。

この言葉は、ストリーミングメディアの発展の段階を三つに分けた上で、最終的には標準化された技術に基づいたストリーミングの段階が訪れると予測したものである。ストリーミングメディアの技術は、今まさにそのような段階に達しているのではないだろうか。

ストリーミングメディアに限られた問題ではないが、インターネット上の（というよりは、おおよそすべてのシステムの）サービスで問題となるのは、サービスの送り手と受け手、それから両者の間でやり取りされるデータの形式の三つが共通の規約に従って実装され、相互運用性が確保されることである。インターネットが爆発的に普及するきっかけとなったのも、サービスの送り手であるサーバ（特にウェブサーバ）、サービスの受け手であるクライアント（特にウェブブラウザ）そして、データ形式には様々な層が介在するものの、もっとも表層的にはHTMLがあったことである。このために、HTMLを適切に処理するソフトウェアである限り、ユーザはブラウザソフトウェアを自由に選ぶことができ、またサービス提供者もサーバを自由に選ぶことができた。

ストリームメディアの標準化が進むためには、もう一つ課題がある。それは、サーバとクライアントが相互にやり取りするためのプロトコルの標準化の問題である。RTSP/RTPなどのストリームメディアのための通信プロトコルが既にIETFなどの場で制定されているが、WWWのように、通信の両端（サーバおよびクライアント）で使用するソフトウェアが任意に選択できるまでには至っていない。

5. おわりにかえて マークアップ言語によるメディアの統合

ストリームメディアの技術は、それ自体が一つのまとまりをもつ単一のメディアのための符号化技術として始まった。しかし、実際のコンテンツは、多様なメディアの組み合わせによって構成される。そのために、単一のメディアを組み合わせるための方式が別に必要となる。

この目的を達成するためのアプローチの一つは、メディアの集まりを、さらに上位のバイナリ符号化によって統合するというアプローチである。例えば、QuickTimeのマルチトラック機能は、複数のメディアを一つの方式で符号化し、一つの大きなメディア（マルチメディア）にするものと言えるだろう。

それとは別のアプローチとして挙げられるのが、マークアップ方式である。マークアップ方式は、各メディアを別の符号化によってあらたなストリームに組み直

⁸ <http://www.techweb.com/se/directlink.cgi?WIR1996102219>

すことはしない。その代わりに、マークアップ言語によって書かれたマークアップ文書によって、それぞれのメディアの相互の時間関係や論理関係を記述する。

究極的には、どちらのアプローチも、メディアの同期関係を何らかの手段によって記述させるのであるから、二つのアプローチを厳密に区別することはできないが、それでも、マークアップ言語には、バイナリ符号化にはない拡張性と柔軟性が備えられている。WWWの記述言語であるHTMLは、オーサリングの際の敷き居が低く、WWWを身近なものにした。今日、我々が、ハイエンドのプロフェッショナルコンテンツから、ローエンドの入門者向けコンテンツまでを目にすることができるのは、オーサリングの敷き居を下げたHTMLの功績である。

World Wide Web Consortium (W3C) は、マークアップ言語として1998年6月に Synchronized Multimedia Integration Language 1.0 (SMIL) を勧告として公開した⁹。SMILは、動画や音声などのメディアを統合し、マルチメディアからなるコンテンツに、「プレゼンテーション」というまとまりを与えるための技術である。SMILは、メディアの上位に位置づけられるため、個々のメディアの転送の方式については関与しない¹⁰。そのため、下回りの技術が更新されても、プレゼンテーションであるコンテンツの記述の枠組みをその都度変更する必要がない。これは、SMILなど上位のマークアップ言語を利用することのメリットである。インターネットストリームメディアの技術は非常に早く更新されるため、過去のコンテンツがそのまま利用できる仕組みは重要である。

また、SMILは、W3CのXML 1.0 Recommendationに基づいて作成されたマークアップ言語であるため、Document Object Model (DOM) やCascading Style Sheet (CSS) など、他のW3C関連技術を援用することができるほか、これらの技術のための開発環境をそのまま利用することができる。このことは、再生クライアント、送信サーバ、オーサリングツールなどの開発を刺激し、競争を生み出す原動力となり、高い品質のソフトウェアを作り出すことにつながるものと期待できる。

上村 圭介 (かみむらけいすけ)

国際大学グローバル・コミュニケーション・センター 研究員

⁹ <http://www.w3.org/TR/REC-smil/> なお、1999年8月3日に新バージョン作成作業の経過報告である、working draft (作業素案) が公開された。URLは、<http://www.w3.org/TR/smil-boston/>

¹⁰ 現在のSMIL、また現在作業中のSMIL "Boston"では、全く関与しないわけではなく、ユーザクライアントのデフォルト言語や、回線速度などのシステム特性によってプレゼンテーションの内容を変えるための記述をすることができる。例えば、ユーザが「日本語」を標準に設定していれば、字幕を日本語に切り替えたり、回線速度が十分確保できなければ、1.5 Mbpsの代わりに、56Kbpsのビデオを再生する、という指定が可能である。