

第1章 MPEG-7 入門

マルチメディア・コンテンツの検索 / MPEG-7 登場の背景 / メタ情報による検索 / メタデータ表現規格

1.1 MPEG-7 はマルチメディア情報の管理・検索規格

MPEG-7と聞くと、読者のみなさんは、映像や音声の新しい圧縮技術と思われる方も多いのではないかと思います。MPEG-7は、画像、映像、音声などのマルチメディア情報を管理・検索するための新しい規格で、従来のMPEG規格とは、目的も内容も異なる標準です。

近年の通信インフラの整備にともない、インターネットでさまざまな情報を利用している方も多いでしょう。インターネット上の情報検索では、すでにテキストベースの検索は広く利用されています。キーワードを入力することで、Web上のサイトを簡単に探し出すことができるようになりました。これは全文検索の技術で、検索の対象が文字コードで表現されており、キーワードも文字コードであるために確実に検索できます。

一方、画像や映像や音声はどうでしょうか？ 写真や映画や音楽で「あの写真を探したい、あの映画のシーンを探したい、あの曲を探したい」と思ったことは誰でもあると思います。しかしこれらの検索は、テキストの場合のようにはいきません。写真を例に考えると、写真には自分のお気に入りの俳優が写っていても実際の画像データには、どこにもその俳優の名前は書かれていないのです。

このように、個人で管理しているマルチメディア・コンテンツだけでなく、現実には、世界中のマルチメディア・コンテンツに対して共通にアクセスし、特定し、さらに管理・検索したいといった要求が急速に高まってきました。こうした要求や問題を解決するために登場したのが、本書でこれから解説するMPEG-7です。

1.2 MPEG-7 登場のいきさつ

MPEG-7は、MPEG(Moving Picture Experts Group)により開発されたISO/IEC国際標準のひとつです。1992年にMPEG-1が、1994年にはMPEG-2が発表され、1998年にはMPEG-4 Version1が、1999年にはMPEG-4 Version2がリリースされてきました。これらは基本的に映像や音声の圧縮に関する国際標準です。MPEG-1とMPEG-2は、現在、VIDEO-CD、MP3、デジタルオーディオ放送(DAB)、DVD-VIDEO、デジタルテレビ放送(DVB、ATSC)やビデオオンデマンドなどに利用され、広く普及している技術のひとつです。また、MPEG-4は、初のマルチメディア統合表現のひとつで、自然映像や合成映像の複合化やインタラクティブ、オブジェクトベースのコード化を実現し、低い転送レートでも高品質の映像や音声の符号化を実現している技術です。

このようにMPEGは、現在広く利用されている映像や音声の圧縮技術を提供してきました。その結果、従来、容量も大きく、データ転送も困難であったマルチメディア・コンテンツが手軽に利用され、広く流通するようになってきました。MPEGは、みずからマルチメディア・コンテンツの洪水を引き起こし、つぎのミッションとして、それを的確に管理する仕組みを提供するようになったわけで、それがこれから紹介するMPEG-7へとつながっているのです。

1.3 マルチメディア・コンテンツのメタ情報による検索

MPEG-7の内容にふれるまえに、マルチメディア・コンテンツの検索についてもう少し詳しく説明をしておきます。これまでも、マルチメディア・コンテンツの検索にはいくつかの試みが行われてきました。ひとつは、ビット表現されているマルチメディア・コンテンツをビットパターンによりマッチングを行う方法です。この方法では、たしかに同じ(完全一致する)ものを見つけだすことは可能ですが、その情報量を考えるとあまり効率の良い作業ではありませんし、1ビットでも異なると検索できません。人が見てもほとんど同じに見えるJPEG画像でも、その圧縮率が異なるとまったく違ったビットパターンになってしまうため、検索することはできません。さらに、こうした手法を用いても、「だれだれが写っている写真を探したい」といった要求はみたすことができません。

そうした中、マルチメディア・コンテンツを効率よく検索する手段として、メタ情報が注目を集めています。これは検索対象とするマルチメディア・コンテンツからその検索対象となる「特徴」をあらかじめ抽出しておき、これらの「特徴データ」を直接の検索対象とすることでマルチメディア・コンテンツの検索に使おうというものです。ここで、それら特徴データは対象のマルチメディア・コンテンツ(これ自身もひとつのデータ)に対するデータであることから、一般にメタ情報(情報への情報データ)と呼ばれています。このようなメタ情報は、対象となるマルチメディア・コンテンツの特徴をコンパクトに表現したものであれば、必ずしもテキスト情報である必要はありません。ここで、メ

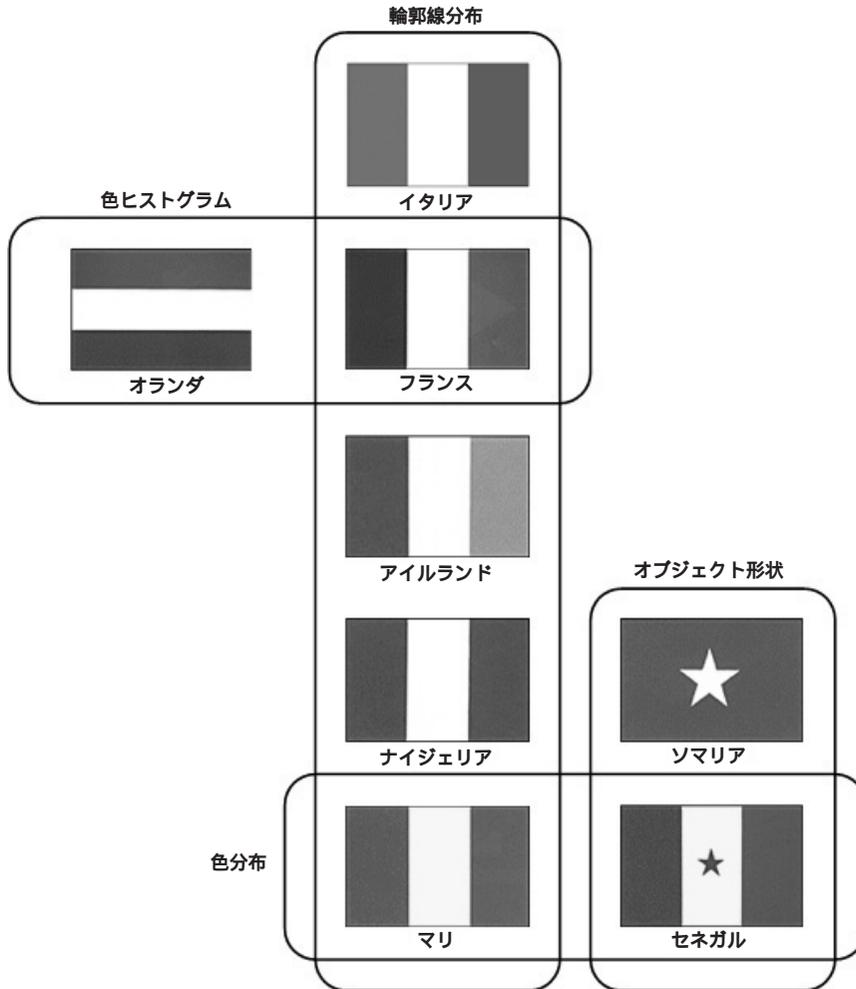


図1.1 特徴量を用いた国旗の分類(カラー口絵参照)

タ情報として表現される特徴量を、色ヒストグラム、輪郭線分布、色分布、オブジェクト形状の4種類に分類してみます。

図1.1をもちいてそれぞれの特徴量とその機能を簡単に紹介しましょう。

色ヒストグラム

色ヒストグラムは、色をいくつかの類似した色のまとまりに分類したものです。簡単にいうと、赤から紫までの色区分を用意し、画像中の各画素がその区分に分類されます。その結果、赤い画素数、青い画素数といったそれぞれの色の分布が抽出され、その分布の類似性を評価することができます。この場合、フランスの国旗とオランダの国旗は、同じような分布を示すので類似とみなすことができます。

輪郭線分布

輪郭線分布では、輪郭線の分布と方向を抽出することで、画像中の色の境界や線の特徴量として表現することができます。その結果、イタリア、フランス、アイルランド、ナイジェリア、マリの国旗は、色の境界が同じような位置に分布することから、類似と判断することができます。

色分布

色分布は、前述の色ヒストグラムと同様にどのような色が存在するかを抽出しますが、同時に色の位置情報も抽出します。その結果、同じ場所に似た色が存在するものが類似とみなされて、マリの国旗とセネガルの国旗が類似と判断されます。

オブジェクト形状

オブジェクト形状は、輪郭線分布をベクトル化することにより、形状を表現します。その結果、位置がちがったり、大きさがちがったりしても同様の形状として判断することができます。さらに、ソマリアの国旗とセネガルの国旗は、同じ星型の形状が存在することで類似と判断されます。

個々の特徴は画像特徴のひとつの側面しか表現することができませんが、このような特徴量をメタデータとして表現しておいて、うまく組み合わせて用いることで、目的の類似した画像を検索することができるようになります。

このように人間が視覚的、聴覚的にとらえる特徴に対しては、これらを言葉で適切に表すことはかえって困難で、むしろ定量的な数値データとして表現したほうがよい場合が多くあります。たとえば、人間の視覚は人によってさまざまで、同じ色でも、人によって「青」と感じたり、「緑」と感じたりといった場合があるからです。そこで、ある共通の色空間を定義しておいて、その色空間での距離を比較するといった方法が有効となります。

このような特徴データは、コンピュータにより自動抽出することが可能ですが、さらに高レベルのメタ情報としては、画像や映像の中身を示す意味的信息があります。

この例では単純ですが、国名をメタ情報として記述しておくことで、簡単にその国の国旗を見つけることができ、さらに国の名前を指定して似た国旗の国を探し出すといったことができるようになるのです。

音声情報検索についても同様のことがいえます。たとえば希望の楽曲を、そのタイトル、アーティスト名などのテキスト情報をキーとして検索するだけでなく、ハミングで奏でた主旋律をキーとして検索することができれば音楽データベースの使い勝手は格段に向上することでしょう。この場合、各音楽データの主旋律を記述したものをあらかじめメタデータとしてデータベース内に保存しておき、他方、入力ハミングから抽出した旋律データをキーにデータベース内のメタデータとのマッチングを行うことで、ミュージック検索システムを実現することができるのです。

また、たとえばDVD等に記録された映画において、ある特定のシーン、たとえば「アクションシ

ーン」だけを見たいという要求がある場合、映画の中のアクションシーンを指し示した「しおり」データをメタデータとしてあらかじめ準備しておけば、それらのシーンへのアクセスは簡単になります。また、アクションシーンのみならず、たとえばある特定俳優のアクションシーンなども全てこのようなメタデータが有効となります。さらにこのような「しおり」データの考えを発展させることにより、本における「目次」や「索引」などに相当するものが、マルチメディア・コンテンツに対しても実現できるのです。

このようにマルチメディア・コンテンツにおけるメタ情報の利用場面を想定すると、とくにオープンなシステムにおいて問題となるのがその表記方法です。つまりデータベース内に蓄積されたメタデータの表記方法と検索キーとして与えるメタデータの表記方法が異なっていれば、前述したようなアプリケーションは実現できません。MPEG-7はこれらマルチメディア・コンテンツのメタデータとして、統一した表記方法を提供する国際規格なのです。

1.4 MPEG-7はメタデータ表現の規格

MPEG-7は、MPEGとして知られるISO/IEC JTC1 SC29/WG11において規格策定されたマルチメディア・コンテンツに対するメタデータの表記方法に関する国際標準規格で、正式名称をMultimedia Content Description Interfaceといいます。MPEG-7は、メタデータすなわちマルチメディア・コンテンツの検索の際に検索対象となる特徴データ(メタデータ)を表現するための規格です。

図1.2にはMPEG-7の標準化対象を模式的に示しています。メタデータを用いたアプリケーションは一般にそれを生成(Generation)する過程とそれを利用(Consumption)する過程に分かれます。双方の間を標準化された表記方法が仲介するという形でモデル化されています。MPEG-7が標準化対象と

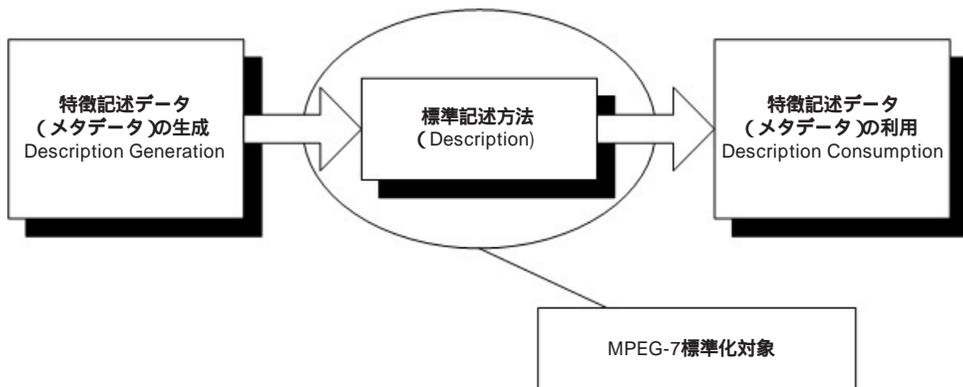


図1.2 MPEG-7の標準化対象

したのはこの標準表記方法であり、MPEG-7は特徴データの抽出方法やその利用法については規定していません。MPEG-7が規格化された時点でも特徴抽出方法には多くの技術的発展の余地があり、応用もさまざまなものが提案されると考えていたからです。

規格が検討されたころから、MPEG-7は特定のアプリケーション形態として、ライブ中継などの実時間 (realtime) 型であるとか、あるいはPull型 (クライアントの要求に応じてサーバがコンテンツを供給する形式) か、Push型 (サーバがクライアントへ一方的にコンテンツを供給する形式) か、といったことにはとらわれず定められた規格であることも特徴のひとつです。またMPEG-7にしたがって記述されたメタデータの取り扱い形態に関してもとくに規定はしていません。たとえばあるアプリケーションにおいては、メタデータは対象となるコンテンツとは別の場所にファイル形式で保存される場合もありますし、また他のアプリケーションではメタデータが対象コンテンツと共に利用者の元へ伝送されるという場合もあり、MPEG-7はその両方の形態に対して利用することが可能です。さらにMPEG-7は、対象となるマルチメディア・コンテンツ自身のメディア形式にも依存していません。つまりDVDやMP3といったデジタル・コンテンツのみならず従来のビデオテープやレコード盤といったアナログ・コンテンツへの適用も可能なのです。

MPEG-7はアプリケーションを限定していません。SMPTE (Society of Motion Picture and Television Engineers)、Dublin Core (Dublin Core Metadata Initiative)、EBU (European Broadcasting Union) P/Meta、TV-Anytime (TV-Anytime Forum) などの標準化団体が策定した特定の用途を想定したメタデータ標準表記規格とは異なります。しかしMPEG-7では、その標準化過程でこれらの標準化団体と積極的なリエゾン関係をもつことで、MPEG-7規格と他のメタデータ標準表記規格との間で相互運用性が保てるよう考慮されています。

このように、MPEG-7はある特定アプリケーション視点のアプローチにしたがったマルチメディア・コンテンツの特徴記述方法を規定しているのではなく、マルチメディア・コンテンツをさまざまな側面からとらえた場合に観察あるいは抽出される特徴の記述を可能とした各種の記述仕様の集合として構成されています。MPEG-7では、こうした記述仕様をツール、それらの集まりをツールセットと呼んでいます。たとえば人間が付加する特徴に関しては、マルチメディア・コンテンツの、タイトル、制作者などの制作関連情報、利用形態、対象コンテンツのメディア形態などの記述ツールセットから、コンテンツの構造やコンテンツの意味内容を記述するためのツールセットなどまでが網羅的に規定されています。

一方、コンピュータ等での処理が可能な信号レベルの特徴に関しては、映像コンテンツがとらえた視覚オブジェクトの色、形状、模様、動きなどを記述するためのツールセットが規定され、また音声コンテンツに対しては、パワー、スペクトラムなどから旋律、さらには音声認識などで用いられる音素の生起確率に関する情報の記述などを可能としたツールが規定されています。このように人間が付加する特徴記述から、信号処理で自動的に得られるような特徴記述までその表記方法が一貫して提供されていることが、MPEG-7の特徴です。規格では前者を高レベル記述、後者を低レベル記述と表現しています。

そして、マルチメディア・コンテンツに関するさまざまな特徴の記述ツールを策定した結果、MPEG-7が規定したツールは膨大な数になりましたが、全て以下の基本要素のいずれかとして分類することができます。MPEG-7が標準化対象としたのは以下に示す5項目です。

記述子(Descriptor : D) : マルチメディア・コンテンツにおいて、ある単一の特徴を記述するための基本ツール。MPEG-7では記述子の表記方法(Syntax)および意味(Semantics)を規定しています。

記述スキーム(Description Scheme : DS) : 複数の記述ツール間の構造、意味的関係を規定した枠組みで、MPEG-7では記述子と同様これら記述スキームの表記方法(Syntax)および意味(Semantics)を規定しています。また、記述スキームを構成する記述ツール要素としては、記述子や下記データ型のほか、他の記述スキーム自体も含まれます。

データ型(Data Type) : 特徴記述ツールの中でもとくにマルチメディアの特徴を記述するものではなく、他の記述子あるいは記述スキームの構成部品として使われるツール。MPEG-7ではデータ型の表記方法(Syntax)および意味(Semantics)を規定しています。

記述定義言語(Description Definition Language : DDL) : 記述子および記述スキームの表記方法を規定するための言語。MPEG-7ではW3C (World Wide Web Consortium)で標準化策定されたスキーマ言語であるXML Schemaをベースに、マルチメディア・コンテンツの特徴記述の際に必要なデータ型などを多少追加することで記述定義言語を規定しています。

システムツール(System Tools) : MPEG-7メタデータを効率的に伝送、蓄積し、またコンテンツとの同期をとるためのアーキテクチャおよびツールを規定しています。

記述定義言語(DLL)にXML Schemaを採用しているため、MPEG-7ではメタデータはXML (eXtensible Markup Language)を用いて表現されます。つまりMPEG-7メタデータは、記述子あるいは記述スキームとして定義された要素および構造にしたがったXML文書として生成(インスタンス*化)されます。実際にはMPEG-7メタデータは<Mpeg7>を開始タグとしたXML文書でリスト1.1のように表現されます。

XMLは、汎用的に利用できる点で情報処理分野において注目を浴び、さまざまなアプリケーションで広く利用されています。そして現在ではXML文書を編集・生成するさまざまなソフトウェアユーティリティが多く開発、提供されています。MPEG-7メタデータに対してはこれらXML汎用ユーティリティを有効に利用できるのも、さらにHTML(Hyper Text Markup Language)やW3Cで策定されたSMIL(Synchronized Multimedia Integration Language)などへの変換や相互利用も可能です。

* XML Schemaに準じて具体的な記述を行ったものをインスタンスと呼んでいる。

リスト1.1 MPEG-7メタデータはXML文書で表現する

```

<?xml version="1.0" encoding="Shift_JIS"?>
<Mpeg7 xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="urn:mpeg:mpeg7:schema:2001"
  xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001">
<Description xsi:type="ContentEntityType">
  <MultimediaContent xsi:type="ImageType" id="image-1">
    <Image>
      ...
    </Image>
  </MultimediaContent>
</Description>
</Mpeg7>

```

つぎに、マルチメディア・コンテンツに対するさまざまな特徴がXML文書として表現された場合、MPEG-7メタデータ自身のデータサイズが膨大になり、蓄積したり、マルチメディア・コンテンツとともに伝送する際の効率が悪くなることが想定されます。MPEG-7ではXMLにもとづく、人が可読なテキストフォーマットに加えてバイナリフォーマット(Binary format for MPEG-7 data : BiM)も規定しています。これはMPEG-7記述ツール表記法の特徴を最大限に活かして開発されたもので、BiMを用いることによってXML文書として表記されたMPEG-7のインスタンスを完全に可逆な状態で10分の1以下のサイズにまで圧縮することができます。さらにこのバイナリフォーマットに対してはデータ内部へのランダムアクセスが可能だけでなく、データ内部の記述を一部分だけ削除したり、あるいは別の内容に書き換えたりするような操作のための方式も規定しています。

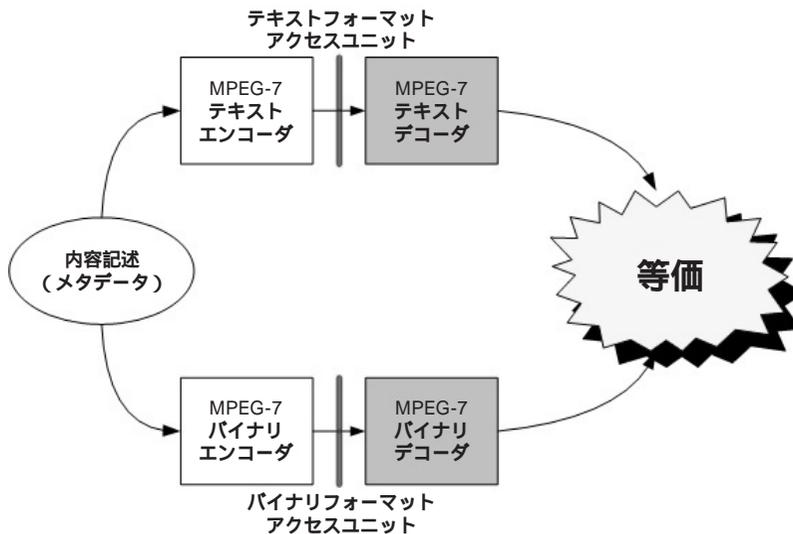


図1.3 MPEG-7における解釈プロセス

図1.3はMPEG-7の解釈プロセスを示しています。MPEG-7ではテキストフォーマットにしたがったメタデータ表記とバイナリフォーマットにしたがったメタデータ表記は情報としては全く同じであり、テキストフォーマットとバイナリフォーマットは相互変換が可能です。その結果XML文書として表記されたMPEG-7メタデータをバイナリ表現して、再びXMLテキスト表現しても元の情報を完全に保つことができます。

今まで述べてきたMPEG-7の主要構成要素の関係を模式的に表したものが図1.4です。

MPEG-7の構成文書

MPEG-7はPart-1からPart-10で構成されています。以下に、MPEG-7の構成をISO/IEC標準文書のパートにしたがって簡単に紹介します。

ISO/IEC 15938-1

・ Part-1 Systems

MPEG-7を利用するアプリケーションの標準的な端末モデルの他、MPEG-7メタデータのアクセス単位に関する規定および先に紹介したMPEG-7メタデータのバイナリフォーマットに関する規定の詳細が記載されています。

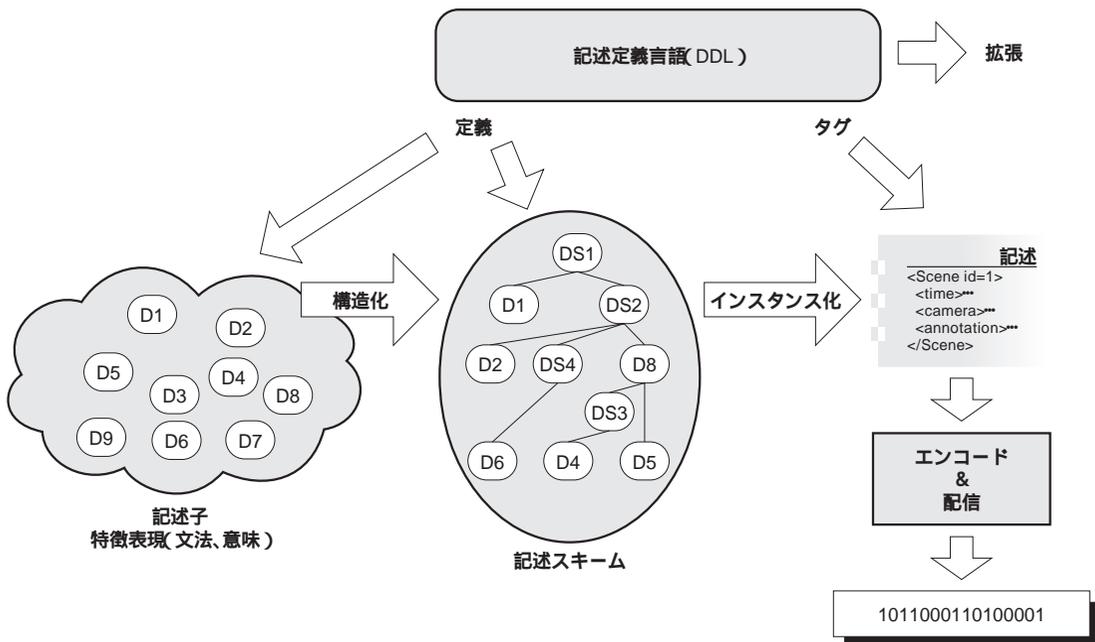


図1.4 MPEG-7の主要構成要素

ISO/IEC 15938-2

・ Part-2 Description Definition Language

W3Cが策定したXML Schemaに対するMPEG-7拡張として、ベクトルおよび行列データ型、時間データ型が追加規定されています。またDDLの基礎となるXML Schema仕様の概要が付録として紹介されています。

ISO/IEC 15938-3

・ Part-3 Visual

映像コンテンツに対する色、形状、模様、動き、画像内における部分指定を記述するためのツール、ビデオなどにおける各特徴データの時間分布を特徴データの並びあるいは補間によって記述可能としたツールなどの記述子(D)および記述スキーム(DS)が規定されています。

ISO/IEC 15938-4

・ Part-4 Audio

音声コンテンツに対するパワー波形、パワー・スペクトラム、ピッチ、音色、無音領域などの低レベル特徴、音声コンテンツの分類に関するツールなどの記述子(D)および記述スキーム(DS)が規定されています。

ISO/IEC 15938-5

・ Part-5 Multimedia Description Scheme

マルチメディア・コンテンツの特徴記述に必要なツールのうち、先のPart-3およびPart-4に直接含まれない全てのツールが本パートで規定されています。実際には以下の6項目に分類し規定しています。

- (1) 基本要素 (Basic Elements)
- (2) コンテンツ管理ツール (Content Management)
- (3) コンテンツ記述ツール (Content Description)
- (4) コンテンツナビゲーション、アクセスツール (Navigation & Access)
- (5) コンテンツ組織化ツール (Content Organization)
- (6) ユーザ関連ツール (User Interaction)

ISO/IEC 15938-6

・ Part-6 Reference Software

上記Part-3からPart-5で規定された各種記述子(D)、記述スキーム(DS)およびPart-1で規定されたMPEG-7バイナリ(BiM)符号化/復号化ツールの実装例が参照ソフトウェアとして紹介されています。