



拡張可能なマーク付け言語 (XML) PMO

W3C勧告 P0日 2月 PXXW年

本版

<http://www.w3.org/TR/PXXW-qdbLxmILXWOQPO>
<http://www.w3.org/TR/PXXW-qdbLxmILXWOQPO.xml>
<http://www.w3.org/TR/PXXW-qdbLxmILXWOQPO.html>
<http://www.w3.org/TR/PXXW-qdbLxmILXWOQPO.pdf>
<http://www.w3.org/TR/PXXW-qdbLxmILXWOQPO.ps>

最新版

<http://www.w3.org/qdbLxmILXWOQPO>

旧版

<http://www.w3.org/old/qdbLxmILXWOQPO>

作成者及び寄稿者

sim a-ay Gsexuality and metescapeH [tb-ay_textuality@com]
iean oaoi G I ic-osofth [jeanpa_mic-osoft@com]
bM I M rpe-be-gl I cpueen G t nive-sity of hllinois at bhicagoH [cmsmcq_uic@ledu]

[copy-right](#) © PXXW年 v Rb® G I hsk hmqhAK j eioHK All qights qese-vedM
本文書には v Rbの[免責](#)、[商標](#)、[文書使用](#)及び[ソフトウェア使用許諾](#)の規定が適用される。

概要

拡張可能なマーク付け言語 (XML) は r f l k のサブセットであって、この標準情報 (XML) で、そのすべてを規定する。w l k の目標は、現在の g s l k と同様に、任意の文書型を持つ r f l k 文書を v v v 上で配布、受信及び処理できることとする。w l k は実装が容易であって、r f l k 及び g s l k のどちらに対しても相互運用性を保つ設計がなされている。

本文書の状態

この標準情報 (XML) の原勧告は v Rb メンバ及び作業グループによってレビューされ、v Rb 勧告として技術統括責任者によって承認された。この標準情報 (XML) は、安定したものであって、参考

資料として使用してもよく、他の文書から規定の参考資料として引用してもよい。勧告作成において、v Rbは、この規定を広く知らせ、普及させる役割をもつ。これによってv v vの機能と相互運用性が高まる。

この標準情報Gs qHは、現在広範囲に使用されている国際的なテキスト処理の標準Grtanda-d f enL e-alized l a-kup kanguageK hrn WWVXPXWUに修正を加えたものHをv v v上で使用するためにサブセット化した構文を規定する。この標準情報Gs qHは、w l k活動を通じて作成されたもので、この活動の詳細は <http://www.w3.org/TR/2006/REC-xsl-fo-20061212/> に示されている。現在のv Rb勧告のリスト及び他の技術文書は<http://www.w3.org/TR/>に示されている。

この標準情報Gs qHでは、[zae-ne-slkee et al](http://www.w3.org/TR/2006/REC-xsl-fo-20061212/)に定義する t qhG t nifo-m qesou-ce hdentifie-Hを使用する。t qhの制定作業は進行中であって、[zhse qebPVRW](http://www.w3.org/TR/2006/REC-xsl-fo-20061212/)及び[zhse qebPWOW](http://www.w3.org/TR/2006/REC-xsl-fo-20061212/)を更新する予定となっている。

この標準情報Gs qHの原勧告の正誤表は <http://www.w3.org/TR/2006/REC-xsl-fo-20061212/>において入手できる。

この標準情報Gs qHの原勧告に誤りがあれば xml-editor@w3.orgに報告されたい。

目次

PI	一般事項	P
PIPI	経緯及び目標	P
PIIQ	定義	Q
QI	文書	R
QIPI	整形式のw I k文書	R
QIQI	文字	S
QIRI	共通の構文構成子	S
QISI	文字データ及びマーク付け	T
QITI	コメント	T
QIU	処理命令	U
QIVI	b c As A セクション	U
QIWI	前書き及び文書型宣言	U
QIXI	スタンドアロン文書宣言	W
QIPO	空白の取扱い	X
QIPP	行末の取扱い	PO
QIPQ	言語識別	PO
RI	論理構造	PP
RIPI	開始タグ，終了タグ及び空要素タグ	PQ
RIQI	要素型宣言	PR
RIQIPI	要素内容	PS
RIQIQI	混合内容	PT
RIRI	属性リスト宣言	PT
RIRIPI	属性の型	PU
RIRIQI	属性のデフォルト	PV
RIIRI	属性値の正規化	PW
RIISI	条件付きセクション	PW
SI	物理構造	PX
SIPI	文字参照及び実体参照	QO
SIQI	実体宣言	QP
SIQIPI	内部実体	QP
SIQIQI	外部実体	QQ
SIRI	解析対象実体	QR
SIRIPI	テキスト宣言	QR
SIRIQI	整形式の解析対象実体	QR
SIRIRI	実体における文字符号化	QR

SISII	w l kプロセサによる実体及び参照の扱い	QS
SISIIPI	“ 認識しない ”	QT
SISIIQI	“ 取込み ”	QT
SISIIRI	“ 検証のために取込み ”	QT
SISISII	“ 禁止 ”	QU
SISITII	リテラル内での取込み	QU
SISIUII	“ 通知 ”	QU
SISIVII	“ 処理しない ”	QU
SISIWII	“ odとして取込み ”	QU
SITII	内部実体置換テキストの構築	QU
SIIUII	定義済み実体	QV
SIVII	記法宣言	QV
SIIVII	文書実体	QW
TII	適合性	QW
TIIPII	妥当性を検証するプロセサ及び検証しないプロセサ	QW
TIIQII	w l kプロセサの使用	QX
UII	表記法	QX
附属書		
AII	参考文献	RO
AIIPII	規定の参考文献	RO
AIIQII	他の参考文献	RP
aII	文字クラス	RQ
bII	w l k及びr f l k【非準】	RS
cII	実体参照及び文字参照の展開【非準】	RT
dII	決定的内容モデル【非準】	RU
eII	文字符号化方式の自動検出【非準】	RU
fII	v Rb w l k作業グループ【非準】	RW

PII 一般事項

拡張可能なマーク付け言語w l kGewtensible l a-kup kanguageHは、w l k文書というデータオブジェクトのクラスを規定し、w l k文書进行处理するプログラムの動作の一部を規定する。w l kは、r f l kGrtanda-d f ene-alized l a-kup kanguageHh r n W W V X |の制限したサブセットとする。w l k文書は、必ずr f l k規格に適合する。

w l k文書は**実体**という記憶単位から成り、実体は構文解析されるデータ又は構文解析されないデータから成る。構文解析されるデータは、**文字**から成り、その一部は**文字データ**を構成し、一部は**マーク付け**を構成する。マーク付けは、文書の記憶レイアウト及び論理構造を記述する符号とする。w l kは、記憶レイアウト及び論理構造についての制約条件を記述する機構を提供する。

w l kプロセサ というソフトウェアモジュールは、w l k文書を読み込み、その内容及び構造へのアクセスを提供するために用いる。 w l kプロセサは、他のモジュールのために動作することを前提としており、そのモジュールをアプリケーションという。この標準情報Gs q Hは、w l kプロセサに要求される振舞いを規定する。つまり、w l kデータの読み込み方法を規定し、アプリケーションに提供する情報を規定する。

PIII 経緯及び目標

P X X U年にv o-l d v i d e v e b b o n s o-t i u m G v R b Hの中に設立されたw l k作業グループG以前は、r f l k編集レビュー委員会と呼ばれた。Hがw l kを開発した。この作業グループの議長を、r u n l i c-o s y s t e m sのi o n a o s a kが務めた。v R bが組織し、以前はr f l k作業グループと呼ばれたw l k r h f G r p e c i a l h n t e-e s t f-o u p Hも、w l kの制定に非常に活発に参画した。w l k作業グループのメンバーを附属書fに示す。c a n b o n n o l l yは、作業グループとv R bとの調整役を務めた。

w l kの設計目標を次に示す。

PII w l kは、インターネット上でそのまま使用できる。

QII w l kは、広範囲のアプリケーションを支援する。

RII w l kは、r f l kと互換性をもつ。

SII w l k文書进行处理するプログラムは容易に書ける。

TII w l kでは、オプションの機能はできるだけ少なくし、理想的には一つも存在しない。

UII w l k文書は、人間にとって読みやすく、十分に理解しやすい。

VII w l kの設計は、すみやかに行う。

WII w l kの設計は、厳密で、しかも簡潔なものとする。

XII w l k文書は、容易に作成できる。

POII w l kでは、マーク付けの数を減らすことは重要ではない。

w l k第PIIO版を理解し、それを処理する計算機プログラムを書くために十分な情報は、この標準情報Gs q H、関連する規格などG文字についてはt n i c o d e及びh r n l h d b P O U S U、言語識別タグについてはh d s e q e b P V U U、言語コードについてはh r n U R X、並びに国コードについてはh r n R P U U。Hによってすべて示す。

この版のw l kの規定は、テキスト及び法律上の注意を一切改変しない限り、自由に配布してもよいIII

PM2M 定義

w l k文書を規定するために使用する用語は、この標準情報Gs qH内で定義する。次に示す語句は、それらの用語を定義するため、及びw l kプロセサの動きを規定するために使用する。

PIIQIP してもよいGmayH

適合する文書又はw l kプロセサは、記述されたとおりに動作してもよいが、そのとおりにする必要はない。

PIIQIQ しなければならないGmustH

適合する文書又はw l kプロセサは、記述されたとおりに動作することが要求される。そうでなければ、エラーとする。

PIIQIR エラーGe-o-H

この標準情報Gs qHが定める規則に対する違反。結果は定義しない。適合するソフトウェアは、エラーを検出して報告してもよく、エラーから回復してもよい。

PIIQIS 致命的エラーGfatal e-o-H

適合するw l kプロセサが検出し、アプリケーションに報告しなければならないエラー。プロセサは、致命的エラーを発見したあとも、それ以降のエラーを探すためにデータ処理を続行し、見つかったエラーをアプリケーションに報告してもよい。エラー訂正をサポートするために、プロセサは、処理していないデータG文字データ及びマーク付けの混在したもの。Hを文書から取り出し、アプリケーションに渡してもよい。しかし、プロセサは、致命的エラーを一度でも検出したなら通常の処理を続行してはならない。つまり、プロセサは、文字データ及び文書の論理構造についての情報を、通常の方法でアプリケーションに渡し続けてはならない。

PIIQIT ユーザのオプション指定によってはGat use- optionH

適合するソフトウェアは、記述されたとおりに振る舞ってもよいGmayH、又は振る舞わなくてはならないGmustHG文章中の助動詞による。H。そのとおりに振る舞う場合は、記述された振舞いを選択又は拒否する手段をユーザに提供しなければならない。

PIIQIU 妥当性制約Gvalidity const-ainth

すべての**妥当な** w l k文書に適用する規則。妥当性制約の違反は、エラーとする。ユーザのオプション指定によっては、**検証を行うw l kプロセサ**は、このエラーを報告しなければならない。

PIIQIV 整形形式制約G•eIlLfo-medness const-ainth

すべての**整形形式**のw l k文書に適用する規則。整形形式制約の違反は**致命的エラー**とする。

PIIQIW マッチGmatchH

aH 文字列又は名前のマッチ 比較する二つの文字列又は名前は、同一でなければならない。hrn\hdb POUSUにおいて、複数の表現が可能な文字 [例えば、合成形式及び基底文字J発音符GダイアクリティカルマークH形式] は、どちらの文字列も同じ表現のときに限り、マッチする。ユーザのオプション指定によっては、プロセサは、その文字を標準形に正規化してもよい。比較のとき、大文字と小文字との区別をする。[aq]bH 文字列と文法中の規則とのマッチ ある生成規則から生成する言語に、ある文字列が属するとき、この文字列は、この生成規則にマッチするという。[aq]cH 内容と内容モデルとのマッチある要素が、制約**A要素の妥当性A**に示す意味で適合するとき、この要素は、その宣言にマッチするという。

PIQIX 互換性のためにはGfo- compatibilityH

w l kの機能であって、w l kがr f l kと互換であることを保証するためだけに導入されるもの。

PIQIPO 相互運用性のためにはGfo- inte-ope-abilityH

拘束力をもたない推奨事項。hrn WWWXへのvebr f l k適用附属書以前から存在するr f l kプロセサが、w l k文書进行处理できる可能性を高めるために取り入れるもの。

2 文書

この標準情報GsqHで定義する意味で、**整形式**のデータオブジェクトをw l k文書という。整形式のw l k文書が、ある制約条件を満足すれば、**妥当なw l k文書**という。

w l k文書は、論理構造及び物理構造をもつ。物理的には、文書は、**実体**という単位からなる。実体が他の実体を**参照**すれば、参照された実体も文書の一部になる。文書は、“ルート”すなわち**文書実体**から始まる。論理的には、文書は、宣言、要素、コメント、文字参照及び処理命令を含み、これらすべては、文書内で明示的なマーク付けによって示す。論理構造及び物理構造は、QRページの § SIIIQ – **整形式の解析対象実体**に示すとおりに、厳密に入れ子になっていなければならない。

2.1 整形式のXML文書

あるテキストオブジェクトが、次の条件を満たすとき、そのテキストオブジェクトを整形式のw l k文書と呼ぶ。

- aH 全体として、**document**というラベルをもつ生成規則にマッチする。
- bH この標準情報GsqHで定義するすべての整形式制約に従う。
- cH 文書内で直接的または間接的に参照されるそれぞれの**解析対象実体**が**整形式**となる。

zP| document W= p-olog element | iscl

document生成規則にマッチするとは、次を意味する。

- aH 一つ以上の**要素**を含む。
- bH ルート又は文書要素という要素が一つだけ存在し、これは、他の要素の**内容**に含まれない。これ以外のすべての要素は、その開始タグが他の要素の内容に含まれれば、対応する終了タグも同じ要素の内容に含まれる。つまり、要素は、開始タグ及び終了タグによって区切られ、入れ子構造をなす。

これらの結果として、文書内のどの非ルート要素Cに対しても、ある他の要素Pが存在し、Cは、Pの内容に含まれるが、Pの内容に含まれる他の要素に含まれることはない。このとき、PをCの親といい、CをPの子という。

2.2.2 文字

解析対象実体は、テキストG文字の並びであって、マーク付け又は文字データを表してもよい。Hを含む。文字は、テキストの最小単位であって、zhrnNhdb POUSUに規定されている。使用できる文字は、タブ、改行、復帰及びGtnicode及びzhrnNhdb POUSUに規定するH図形文字とする。zt nicode|のUMW節で定義される互換性文字は使用を避けることが望ましい。

zQ| bha- Y= BxX | BxA | Bxc | zBxQ0LBxcVee| | zBxd000LBxeeec | | NI 任意の t nicode文字。ただし、サロゲートブロック、eeed及びeeeeは除く。 IN

文字番号をビットパターンに符号化する機構は、実体ごとに違ってもよい。すべてのwIkプロセサは、zhrnNhdb POUSUのtselW符号化方式及びtselPU符号化方式を受け付けなければならない。二つのどちらが用いられているかを明示するための機構、及び他の符号化方式を利用するための機構は、QRページの§SIRIR – 実体における文字符号化に記述する。

2.2.3 共通の構文構成子


QIRでは、文法内で広く使用するいくつかの記号を定義する。

r G空白Hは、一つ以上のスペース文字GBxQOH、復帰、改行又はタブから成る。

zR| r Y= GBxQO | BxX | Bxc | BxAHJ

便宜上、文字を、字、数字又は他の文字に分類する。字は、アルファベットの、若しくは表音的である基底文字G一つ以上の結合文字が、後に続くこともある。H、又は統合漢字から成る。各クラスにおける実際の文字についての完全な定義は、RQページの附属書 a 文字クラスに示す。

mameは、字又はいくつかの区切り文字の一つで始まり、その後に字、数字、ハイフン、下線、コロン又はピリオドが続く。これらの文字を名前文字という。文字列Axm1Aで始まる名前、又は正規表現(('X' | 'x') ('M' | 'm') ('L' | 'l'))にマッチする任意の文字列で始まる名前は、この標準情報GsQHの現在の版又は将来の版での標準化のために予約する。

 wIkの名前の中のコロンは、名前空間での実験のために予約する。コロンの意味は、将来のある時点で標準化するものとし、そのときには、実験的な目的でコロンを使用する文書を更新する必要性が生じる可能性がある。wIkで採用する名前空間の機構が、区切り子として実際にコロンを使用するという保証はない。事実上、これは、名前空間の実験の一つとして以外には、wIkの名前の中でコロンを使用しないほうがよいことを意味する。しかし、wIkプロセサは、名前文字としてコロンを受け付けることが望ましい。

mmtokenG名前トークンHは、名前文字の列とする。

zS| mamebha- Y= kette- | cigit | RIF | FLF | F_F | FYF | bombingbha- | dxtende-

zT| mame Y= Gkette- | F_F | RYFH Gmamebha-HI

zU| mames Y= mame Gr mameHI

zV| mmtoken Y= Gmamebha-HJ

zW| mmtokens Y= mmtoken Gr mmtokenHI

リテラルデータは、引用符で囲まれた文字列とし、その列の区切り子として使用する引用符は含まない。リテラルは、内部実体`<entityvalue>`、属性値`<attributevalue>`、外部識別子`<systemkite-al>`の内容の指定に使用する。`<systemkite-al>`はマーク付けの走査を行わずに解析できることに注意せよ。

```

zX|      entityvalue  Y=  FAF Gz^DEA| | odqefe-ence | qefe-enceHI FAF
          | AFA Gz^DEf| | odqefe-ence | qefe-enceHI AFA

zPO|      Attribute  Y=  FAF Gz^[ EA| | qefe-enceHI FAF
          | AFA Gz^[ Ef| | qefe-enceHI AFA

zPP|      systemkite-al Y=  GFAF z^A| | FAFH | GAz^f| | AFAH

zPQ|      oubidkite-al Y=  FAF oubidbha-| FAF | AFA Goubidbha- L AFAHI AFA

zPR|      oubidbha-  Y=  BxQO | Bxc | BxA | zaLzALyOLX| | zLFghJKlMnY=^Z@| B_ C_ D|

```

2MSM 文字データ及びマーク付け

テキストは、文字データ及びマーク付けから成る。マーク付けは、開始タグ、終了タグ、空要素タグ、実体参照、文字参照、コメント、bcAsAセクションの区切り子、文書型宣言及び処理命令の形をとる。

マーク付けではないすべてのテキストは、文書の文字データを構成する。

アンド記号`&`及び不等号`<`小なり`<` `>`は、マーク付けの区切り子として、又はコメント、処理命令若しくはbcAsAセクション内で使用する場合にだけ、そのままの形で出現してよい。これらの文字は、内部実体宣言のリテラル実体値内に記述してもよい。詳しくは、QRページの§SIRIQ-整形形式の解析対象実体を参照。これらの文字が他の部分で必要な場合、番号による文字参照又は文字列`
`及び文字列`
`を使用して別扱いしなければならない。不等号`>`大なり`>` `>`は、文字列`>A`を使用して表現してもよい。内容の中で列`A]]>A`を使用するときは、それが、bcAsAセクションの終了をマーク付けしない限り、互換性のため、`>A`又は文字参照を使用して別扱いしなければならない。

要素の内容では、文字データは、いかなるマーク付けの開始区切り子を含まない任意の文字列とする。bcAsAセクションでは、文字データとは、bcAsAセクションの終了区切り子`A]]>A`を含まない任意の文字列とする。

属性値が一重引用符及び二重引用符を含むためには、アポストロフィ又は一重引用符`'`は、`'`として表現し、二重引用符`"`は、`"`として表現する。

```

zPS|      bha-cata  Y=  z^[E| | L Gz^[E| | f]]f z^[E| | H

```

2MTM コメント

コメントは、他のマーク付けの外ならば、文書のどこに現れてもよい。さらに、文書型宣言の中で、文法が許す場所に現れてもよい。コメントは、文書の文字データの一部ではない。w l kプロセサは、アプリケーションがコメントのテキストを取り出すことを可能としてもよいが、そうしなくともよい。互換性のためには、文字列 `--A` (二連ハイフン) は、コメント内で現れてはならない。

```

zPT|      bomment  Y=  F[ @LLF GGbha- L FLFH | GFLF Gbha- L FLFHfH | FLL ]F

```

コメントの例を次に示す。

```
<!-- declarations for <head> & <body> -->
```

2.6.6 処理命令

処理命令`<oh>`によって、アプリケーションのための命令を文書に入れることができる。

```
zPU|          oh Y= F[^\ ohsa-get Gr Gbha-| L Gbha-| ^]F bha-|HHH^ F^]F
zPV|          ohsa-get Y= mame L GGfWf | FxRH GF | F | FmRH GFkF | F|FHH
```

`oh`は、文書の文字データの一部ではないが、アプリケーションに渡されなければならない。`oh`は、命令が渡されるアプリケーションを特定するために使用するターゲット`<ohsa-get>`で始まる。ターゲット名`AXMLA`、`AxmlA`などは、この標準情報`GsqH`の現在の版又は将来の版の標準化のために予約する。`wlk`の記法機構を、`oh`のターゲットを宣言するために使用してもよい。

2.6.7 CDATAセクション

`<bcAsA>`セクションは、文字データが出現するところであれば、どこに出現してもよい。これは、`<bcAsA>`セクションで囲まなければマーク付けとして認識されてしまう文字を含むテキストを別扱いするのに使用する。`<bcAsA>`セクションは、文字列`A<![CDATA[A`で始まり、文字列`A]]>`で終わる。

```
zPW|          b c r e c t Y= b c r t a - t b c a t a b c d n d
zPX|          b c r t a - t Y= F[ @z b c A s A z
zQO|          b c a t a Y= G b h a - | L G b h a - | F | ] F b h a - | H H
zQP|          b c d n d Y= F | ] F
```

`<bcAsA>`セクション内では、文字列`<bc dnd>`だけをマーク付けとして認識するので、不等号`<`小なり`H`及びアンド記号は、そのままの形で出現してよい。`A<`、`A`及び`A&`、`A`を使用して別扱いする必要はない。`<bcAsA>`セクションは入れ子にはできない。

`<greeting>``A` 及び`</greeting>``A`を、マーク付けではなく、文字データとして認識する`<bcAsA>`セクションの例を次に示す。

```
<![CDATA[<greeting>Hello, world!</greeting>]]>
```

2.6.8 前書き及び文書型宣言

`wlk`文書は、使用する`wlk`の版を指定する`wlk`宣言で始めてもよく、又そうするのが望ましい。例えば、次に示す完全な`wlk`文書は、整形形式であるが妥当ではない。

```
<?xml version="1.0"?>
<greeting>Hello, world!</greeting>
```

次の文書も同様とする。

```
<greeting>Hello, world!</greeting>
```

この標準情報`GsqH`のこの版に適合することを示すためには、版番号`A1.0A`を使用しなければならない。ある文書が、この標準情報`GsqH`のこの版に適合しないとき、値`A1.0A`を使用するのは、エラーとする。この標準情報`GsqH`の今後の版に`A1.0A`以外の値を付与することが、`wlk`作業グ

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

【整形制約】内部サブセット内のパラメタ実体

cscの内部サブセットでは、**パラメタ実体参照**は、マーク付け宣言が出現可能な場所だけに出現できる。マーク付け宣言の一部としては出現できない。この制約は、外部パラメタ実体又は外部サブセットでの参照には適用しない。

内部サブセットのときと同様に、外部サブセットと、cscにおいて参照する任意の外部パラメタ実体とは、非終端記号`ma-kupdecl`によって許される型の一連の完全なマーク付け宣言で構成されなければならない。マーク付け宣言の間には、空白又は**パラメタ実体参照**を置いてもよい。外部サブセット又は外部パラメタ実体の内容の一部は、**条件付きセクション**を使用して無視してもよいが、内部サブセットではこれは許されない。

```
zR0|      extrubset  Y=  sextcecl^extrubsetcecl
zRP|      extrubsetcecl  Y=  G ma-kupdecl | conditional rect | odqefe-ence | r HI
```

外部サブセット及び外部パラメタ実体は、その中では、**パラメタ実体参照**がマーク付け宣言の間だけでなく、マーク付け宣言の内でも認められる、という点でも内部サブセットとは異なる。

文書型宣言付きのw l k文書の例を、次に示す。

```
<?xml version="1.0"?>
<!DOCTYPE greeting SYSTEM "hello.dtd">
<greeting>Hello, world!</greeting>
```

システム識別子 `Ahello.dtdA`が、文書のcscのt qhとなる。

次の例の通り、宣言を局所的に与えることもできる。

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE greeting [
  <!ELEMENT greeting (#PCDATA)>
]>
<greeting>Hello, world!</greeting>
```

外部サブセット及び内部サブセットの両方を使用するときは、内部サブセットが外部サブセットより先に出現したと見なす。これは、内部サブセットの実体及び属性リスト宣言が、外部サブセットの実体及び属性リスト宣言に優先するという効果をもたらす。

2XIII スタンドアロン文書宣言

w l kプロセサは、アプリケーションに文書の内容を渡すが、マーク付け宣言は、この内容に影響を与えることがある。例えば、属性のデフォルト値及び実体宣言は影響を与える。スタンドアロン文書宣言は、w l k宣言の一部分として出現することができ、影響を与えるマーク付け宣言が**文書実体**の外部に出現するかどうかを示す。

```
zRQ|      r ccecl  Y=  r fstandalonef dq GGAF A Gfyesf | fnoFH AFAH | GAF Gfyesf | fnoFH FAHH
```

スタンドアロン文書宣言においては`KAYesA`の値は、**文書実体**の外部に（cscの外部サブセット内に、又は内部サブセットから参照される外部パラメタ実体内に）、w l kプロセサからアプリケーションへと渡される情報に影響するマーク付け宣言が存在しないことを意味する。`AnoA`の値は、その外部マーク付け宣言が存在するか、又は存在する可能性があることを意味する。ス

拡張可能なマーク付け言語 Gw l kH PIO

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

値AdefaultAは、アプリケーションのデフォルトの空白処理モードを、その要素に適用可能とすることを意味する。値ApreserveAは、アプリケーションがすべての空白を保存することを意味する。この宣言の意図は、Axml:spaceA属性の別の指定で上書きしない限り、要素の内容に現れるすべての要素に適用すると解釈する。

文書のルート要素については、この属性の値を指定するか、又はこの属性のデフォルト値がある場合を除いては、アプリケーションによる空白の取扱いについて、いかなる意図も示さないと解釈する。

2P2P 行末の取扱い

w l kの構文解析対象実体は、通常コンピュータのファイル内に保存され、編集の便宜のために複数の行に分けることが多い。これらの行は、普通は、ca-iageL-etu-nGBxcHコード及びlineLfeedGBxAHコードの何らかの組合せによって分けられる。

アプリケーションの処理を簡単にするため、外部解析対象実体又は内部解析対象実体のリテラル実体値が、ABxcBxAAの2文字の連続とするリテラル又はBxcの単独のリテラルを含む場合に、w l kプロセサは、アプリケーションに単一の文字BxAだけを渡さなければならない。この処理は、入力内に存在する改行コードを構文解析の前に正規化することによって、容易に実現できる。H。

2P2M 言語識別

文書処理においては、その文書の中身がどんな自然言語又は形式言語で書かれているか明示することが、役に立つことが多い。w l k文書内の要素のもつ内容又は属性値において使用する言語を指定するために、Axml:langAという名前の特別な属性を、文書内に挿入してもよい。妥当な文書においてこの属性を使用する場合は、他の属性と同様に宣言されなくてはならない。属性の値は、zhds qeb PVUU | “ qebPVUU: 言語識別のためのタグ ” によって規定される言語識別コードに従う。

- zRR| kanguagehc Y= kangcode GFLF rubcodeHI
- zRS| kangcode Y= hr nURXbode | hanabode | t se-bode
- zRT| hr nURXbode Y= GzaLz | zALy|H GzaLz | zALy|H
- zRU| hanabode Y= GfiF | fhFH FLF GzaLz | zALy|HJ
- zRV| t se-bode Y= Gxf | fwFH FLF GzaLz | zALy|HJ
- zRW| rubcode Y= GzaLz | zALy|HJ

kangcodeは、次のどれでもよい。

aH zhr n URX | “ 言語の名前表現のためのコード ” で規定されるQ文字の言語コード

bH hnte-net Assigned numbe-s Autho-ity zhAmA | で登録されている言語コード。これは、先頭がAi-A G又はAI-ALAHで始まる。

cH ユーザによって定められた言語コード、又は私的な使用のために複数の団体間が取り決めたコード。これらは、今後hAmAにおいて標準化又は登録されるコードとの競合を避けるために、先頭をAx-A又はAx-Aで始める。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

【整形形式制約】要素型のマッチ

要素の終了タグの**名前**は、その要素の開始タグにおける要素型（の名前）とマッチしなければならない。

【妥当性制約】要素の妥当性

要素が妥当とは、その要素型（の名前）とマッチする**mame**をもつ宣言（**elementdecl**にマッチするもの）が存在し、さらに次のいずれかの条件を満たす場合とする。

PII aH 宣言が**EMPTY**にマッチし、要素が**内容**をもたない。

QII bH 宣言が**child-en**にマッチし、要素の**子要素**の並びが、内容モデル中の正規表現によって生成される言語に属する。子要素の間に空白（非終端記号**r**にマッチする文字の並び）があってもよい。

RII cH 宣言が **fixed**にマッチし、要素の内容が**文字データ**及び**子要素**からなる。子要素の要素型は、要素の内容モデルに出現する**名前**にマッチする。

SII dH 宣言が**ANY**にマッチし、どの**子要素**の要素型も宣言されている。

3PII 開始タグ、終了タグ及び空要素タグ

空でない任意のw | k要素の始まりは、開始タグによってマーク付けする。

zSO| r sag Y|= f[f mame Gr Att-ibuteH| r^f]f

zSP| Att-ibute Y|= mame dq Attualue

開始タグ及び終了タグ内の**mame**は、要素の 型を表わす。**mame**及び**Attualue**の対を要素の属性指定といい、個々の対における**mame**を属性名といい、**Attualue**の内容G区切り子' 又は"の間のテキストHを属性値という。

【整形形式制約】属性指定の一意性

開始タグ又は空要素タグでは、同一の属性名が二回以上出現してはならない。

【妥当性制約】属性値の型

属性は宣言されていなければならない。属性値の型は、その属性に対して宣言した型でなければならないG属性の型については、PTページの § RII R – 属性リスト宣言を参照。H。

【整形形式制約】外部実体への参照がないこと

属性値には、外部実体への直接的又は間接的な参照を含むことはできない。

【整形形式制約】属性値に<を含まないこと

属性値内で直接的又は間接的に参照する実体GA<t;Aを除く。Hの**置換テキスト**には、<を含んではならない。

開始タグの例を次に示す。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

```
<!ELEMENT br EMPTY>
<!ELEMENT p (#PCDATA|emph)* >
<!ELEMENT %name.para; %content.para; >
<!ELEMENT container ANY>
```

3M2MPII 要素内容

ある型の要素が子要素だけを必ず含みG文字データを含まない。H, それらの間には空白(非終端記号rにマッチする文字)だけしか現れないとき, その要素型は, 要素内容をもつという。この場合, 内容モデルが制約となる。内容モデルは, 子要素の型及び子要素の出現順序を制御する簡単な文法とする。この文法は, 内容素子GcpHから成る。内容素子は, 名前, 内容素子の選択リスト又は内容素子の列リストから構成される。

```
zSV|      child-en  Y=  Gchoice | seqH GF^ | FIF | FJH^
zSW|      cp       Y=  Gmame | choice | seqH GF^ | FIF | FJH^
zSX|      choice  Y=  FGF r^ cp G r^ Ff r^ cp HI r^ HF
zTO|      seq     Y=  FGF r^ cp G r^ HF r^ cp HI r^ HF
```

ここで, mameは, 子として出現してよい要素の型を示す。この文法で選択リストが現れる位置では, 選択リスト内のいずれの内容素子も要素内容の中に現れてよい。列リストに現れる内容素子は, リストで指定する順番のとおり, 要素内容に現れなければならない。名前又はリストの後に出現するオプションの文字は, リスト内の要素又は内容素子が, P回以上任意の回数G+H, 0回以上任意の回数G*H又は0回若しくはP回G?H出現可能なことを規定する。この演算子がない場合は要素又は内容素子が正確に1度だけ現われなくてはならないことを意味する。ここで示す構文及び意味は, この標準情報GsQHにおける生成規則で用いるものと同じとする。

要素の内容が内容モデルにマッチするのは, 列, 選択及び繰返し演算子に従って, 内容の中の要素と内容モデル内の要素型とをマッチさせながら, 内容モデル内の一つのパスをたどれるときに限る。互換性のため, 文書内の要素が, 内容モデルにおける要素型の複数の出現位置とマッチすることは, エラーとする。詳細な規定については, RUページの附属書 4 決定的内容モデルを参照。

【妥当性制約】グループ及びパラメタ実体が厳密な入れ子をなしていること

パラメタ実体の置換テキストは, かっこで囲まれたグループによって, 厳密な入れ子を構成しなければならない。つまり, 選択, 列又は混在部品に, 開きかっこ又は閉じかっこのいずれか一方がパラメタ実体の置換テキストに含まれれば, 他方も同じ置換テキストに含まなければならない。

相互運用性のためには, パラメタ実体参照が 選択, 列又は 混在内容内容に含まれれば, その置換テキストは空でないことが望ましく, 置換テキストの先頭及び末尾の空白でない文字は, コネクタG|又は,Hでない方がよい。

要素内容モデルのいくつかの例を次に示す。

```
<!ELEMENT spec (front, body, back?)>
<!ELEMENT div1 (head, (p | list | note)*, div2*)>
<!ELEMENT dictionary-body (%div.mix; | %dict.mix;)*>
```

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

3IMP 属性の型

wlkの属性の型は、3種類とする。これらは、文字列型、トークン化型及び列挙型とする。文字列型は、値として任意のリテラル文字列をとる。トークン化型は、字句及び意味に関して、次に示す様々な制約をもつ。

```
zTS|      Attstype  Y=  rt-ingsype | sokenizedsype | dnume-atedsype
zTT|      rt-ingsype Y=  fbcAsAf
zTU|      sokenizedsype Y=  fhcf
                               | fhcqdef
                               | fhcqderf
                               | fdmshsxf
                               | fdmshshdrf
                               | fmlsnjdmf
                               | fmlsnjdmrf
```

【妥当性制約】hc

ID型の値は、生成規則Nameにマッチしなければならない。一つのwlk文書内では、一つの名前が、この型の値として複数回現れてはならない。つまり、hcの値は、要素を一意に特定しなければならない。

【妥当性制約】P要素ごとに一つのhc

要素型は、複数のhc属性をもってはならない。

【妥当性制約】hc属性のデフォルト

hc属性は、デフォルトとして、#IMPLIED又は#REQUIREDを宣言しなければならない。

【妥当性制約】hcqde

IDREF型の値は、生成規則mameにマッチしなければならない。IDREFS型の値は、mamesにマッチしなければならない。各々のmameは、wlk文書内に存在する要素のhc属性の値とマッチしなければならない。つまり、IDREFの値は、あるhc属性の値とマッチしなければならない。

【妥当性制約】実体名

ENTITY型の値は、mame生成規則にマッチしなければならない。ENTITIES型の値は、mamesにマッチしなければならない。各々のmameは、cscで宣言する解析対象外実体とマッチしなければならない。

【妥当性制約】名前トークン

NMTOKEN型の値は、nmtoken生成規則にマッチしなければならない。NMTOKENS型の値は、にマッチしなければならない。

列挙型の属性は、宣言した幾つかの値の一つを取ることができる。列挙型には、Q種類ある。

```
zTV|      dnume-atedsype Y=  motationsype | dnume-ation
```

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

```

<!ATTLIST termdef
    id      ID      #REQUIRED
    name    CDATA   #IMPLIED>
<!ATTLIST list
    type    (bullets|ordered|glossary)  "ordered">
<!ATTLIST form
    method  CDATA   #FIXED "POST">

```

3.3.3 属性値の正規化

XMLプロセッサは、属性値をアプリケーションに渡す前、または、妥当性を判定する前に、次のとおり正規化しなければならない。

注①これは原文にはない | 以下の箇条書きは、AとAの間（もしくはfとfの間）の文字の並びに対して繰り返し実行される条件分岐である。一度実行されるたびに、正規化された文字列が先頭から少しずつ構築されていく。

注②これは原文にはない | QMPPで述べたように、文字を読み込むルーチンがbqJkeやbqkeに置き換えているものと考えると理解しやすい。

- 属性値の中の文字参照は、参照される文字で置き換えることによって処理する。
- 実体参照は、実体の置換テキストを再帰的に処理する。
- 属性値の中のスペース文字BxQ0 Bxc BxA BxHは、BxQ0で置き換えることによって処理する。ただし、外部解析対象実体の一部、または、内部解析対象実体のリテラルリテラル実体値の中にbqBxcとkeBxAとが連続して現れる場合は、一つのBxQ0で置き換える。
- 他の文字は、正規化した値の中にそのまま置く。

さらに、属性の型がbcAsAでない場合は、XMLプロセッサは正規化された属性値に対して、次の処理をしなければならない。まず、先頭または末尾にあるスペース文字BxQ0Hをすべて取り除く。つぎに、連続するスペース文字BxQ0Hを一つのスペース文字BxQ0Hに置き換える。

妥当性を検証しないパーサは、宣言が見つからない属性は、すべて、CDATAを宣言しているとして扱うものとする。

3.3.4 条件付きセクション

条件付きセクションとは、**文書型宣言の外部サブセット**の一部であって、制御キーワードの指定によって、cscの論理構造に含めたり、除いたりする部分とする。

```

zUP| conditional rect Y= include rect | igno-erect
zUQ| include rect Y= F[ @zf r^ rhmbk t cdf r^ fz extrubsetcecl f|| ]f
zUR| igno-erect Y= F[ @zf r^ rh f mnqdf r^ fz igno-erectbontents l f|| ]f
zUS| igno-erectbonL Y= hgno-e Gf[ @zf igno-erectbontents f|| ]f hgno-eHl
    tents
zUT| hgno-e Y= bha- l L Gbha- l Gf[ @zf | f|| ]H bha- lH

```

条件付きセクションは、cscの内部サブセット及び外部サブセットと同様に、完全な宣言、コメント、処理命令又は入れ子になった条件付きセクションを、いくつか含んでよい。これらの間に、空白が現れてもよい。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

SMPM 文字参照及び実体参照

文字参照は、hrnNhdb POUSU文字集合の特定の文字，例えば，入力機器から直接入力不可能な文字を参照する。

```
zUU|          bha-qef  Y|=  FEBf zOLX|J  FZ
                | FEBXi zOLXAlfALe|J  FZ
```

【整形形式制約】使用できる文字

文字参照で参照する文字は、の生成規則にマッチしなければならない。

文字参照が A
 で始めれば、終端の ; までの数字及び字は、hrnNhdb POUSUの文字コード位置のPU進数表現とする。文字が A&#A で始めれば、終端の ; までの数字は、文字コード位置のPO進数表現とする。

実体参照は、名前の付いた実体の内容を参照する。一般解析対象実体への参照は、アンド記号 G&H及びセミコロン記号G;Hを区切り子として用いる。パラメタ実体への参照は、パーセント記号G%H及びセミコロンG;Hを区切り子として用いる。

```
zUV|          qefe-ence  Y|=  dntityqef | bha-qef
zUW|          dntityqef  Y|=  FEf mame  FZ
zUX|          odqefe-ence  Y|=  FDf mame  FZ
```

【整形形式制約】実体が宣言されていること

cscをもたない文書，パラメタ実体参照を含まない内部cscサブセットだけをもつ文書，又は Astandalone='yes'A をもつ文書において、実体参照で用いる `mame` は、ある実体宣言に含まれる名前とマッチしなければならない。ただし、整形形式の文書は、実体 `amp` `lt` `gt` `apos` `quot` を宣言する必要はない。パラメタ実体の場合は、宣言は、参照に先行しなければならない。同様に、一般実体の場合は、属性リスト宣言のデフォルト値内での参照より先に、宣言が現れなければならない。

外部サブセット又は外部パラメタ実体で実体を宣言するとき、妥当性を検証しないプロセサが、宣言を読み、処理することを義務づけられないことに注意。それらの文書では、実体は宣言されなければならないという規則は、`standalone=fyesf` の場合のみ、整形形式制約となる。

【妥当性制約】実体が宣言されていること

外部サブセット又は外部パラメタ実体をもっていて、Astandalone='no'Aをもつ文書において、実体参照で用いる `mame` は、ある実体宣言に含まれる名前とマッチしなければならない。相互運用性のためには、妥当な文書はQVページの § SMPM – 定義済み実体で指定した書式によって、実体 `amp` `lt` `gt` `apos` `quot` を宣言することが望ましい。パラメタ実体の場合は、宣言は、参照に先行しなければならない。同様に、一般実体の場合は、属性リスト宣言のデフォルト値内での参照よりも先に、宣言が現れなければならない。

【整形形式制約】解析対象実体

実体参照は、解析対象外実体の名前を含んでいてはならない。解析対象外実体は、ENTITY型又はENTITIES型として宣言した属性値としてだけ参照できる。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

5.2.2 外部実体

内部実体でない実体は外部実体であって、次のとおりに宣言する。

```
zVT|      dxte-nalhc  Y= frxrsd lf r rystemkite-al
          |fot akhb f r oubidkite-al r rystemkite-al
```

```
zVU|      mcatacecl  Y= r fmcAsAf r mame
```

`mcatacecl` が存在すれば、この実体は、一般解析対象外実体とし、そうでなければ、解析対象実体とする。

【妥当性制約】記法が宣言されていること

`mame` は、宣言した記法の名前とマッチしなければならない。

`rystemkite-al` を、実体のシステム識別子と呼ぶ。システム識別子としては `tqh` を使い、その実体の内容を取り出すために用いてもよい。`tqh` と共に使うことの多いシャープ記号 `G##H` 及びフラグメント識別子は、正式には、`tqh` 自体の一部ではない。フラグメント識別子が、システム識別子の部分として与えられている場合、`wlk` プロセサは、エラーを出してもよい。この標準情報 `GsqH` の適用範囲外の情報 `G` 例えば、ある特定の `csc` の特別な `wlk` 要素又は特定のアプリケーションの仕様によって定義された処理命令 `H` によって上書きされない限り、相対的な `tqh` は、その実体の位置、すなわち、その実体の宣言があるファイルに相対的とする。したがって、その `tqh` は、文書実体、外部 `csc` サブセットを含む実体、又は、いくつかの外部パラメタ実体に対して、相対的である。

`wlk` プロセサは、非 `Arbhh` 文字が `tqh` に含まれている場合、これを次のとおりに扱う。非 `Arbhh` 文字は、`tseLW` によって一つ以上のバイトで表現し、これらのバイトを `tqh` の別扱い機構を用いて（すなわち、バイト値のPU進数による表現を `gg` としたとき、各バイトを `Dgg` の形式に変換することによって）別扱いする。

z訳注 `G`（これは原文にはない） | この規定は `gs | k SIO` の附属書 `a` と `qeb QPSPA t qm ryntaxAK q|| | oatsk | ay PXXV` にあるものである。

システム識別子以外に、外部実体は、公開識別子を含んでもよい。実体の内容を取り出す `wlk` プロセサは、この公開識別子を用いて、代替りの `tqh` の生成を試みてもよい。`wlk` プロセサがこれに失敗した場合は、システムリテラルとして指定した `tqh` を用いなければならない。マッチする前に、公開識別子内にある空白文字からなる文字列は、すべて単一のスペース文字 `GBxQOH` に正規化しなければならない。先頭及び末尾の空白文字はすべて削除しなければならない。

外部実体宣言の例を、次に示す。

```
<!ENTITY open-hatch
  SYSTEM "http://www.textuality.com/boilerplate/OpenHatch.xml">
<!ENTITY open-hatch
  PUBLIC "-//Textuality//TEXT Standard open-hatch boilerplate//EN"
  "http://www.textuality.com/boilerplate/OpenHatch.xml">
<!ENTITY hatch-pic
  SYSTEM "../grafix/OpenHatch.gif"
  NDATA gif >
```

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

符号化宣言では、値`AUTF-8A`、`AUTF-16A`、`AISO-10646-UCS-2A`及び`AISO-10646-UCS-4A`は、`tnicode`及び`hrnNhdbPOUSU`の各種符号化方式のために用いる。値`AISO-8859-1A`から`AISO-8859-9A`までは、`hrn WWTX`の対応する部分のために用いる。値`AISO-2022-JPA`、`AShift_JISA`及び`AEUC-JPA`は、`ihr wLOQOWLPXXV`の各種符号化方式のために用いる。`wlk`プロセサは、ここに挙げた以外の符号化方式を認識してもよい。`hnte-net Assigned mumber-s Autho-ity zhAmA`に、`Gcha-sets`として登録された文字符号化方式については、ここに挙げたもの以外についても、登録された名前で参照することが望ましい。これらの登録された名前は、大文字・小文字の区別をせずに定義されているので、これらに対する比較を試みるプロセサは、大文字・小文字の区別をしない方法をとることに注意する。

外部の伝送プロトコル（すなわち、`gssok lhd`など）で与えられる情報が存在しないとき、`wlk`プロセサに渡された実体が、符号化宣言を含むにもかかわらず、宣言で示したものの以外の方式で符号化されている場合、符号化宣言が外部実体の最初以外の位置に出現した場合、又はバイト順マークでも符号化宣言でも始まらない実体が、`t selW`以外の符号化方式を使用した場合は、**エラー**とする。`Arbhh`は`t selW`のサブセットなので、通常の`Arbhh`の実体は厳密には符号化宣言を必要としないことに注意。

処理できない符号化方式を使用した実体を`wlk`プロセサが発見したときは、**致命的エラー**とする。

符号化宣言の例を次に示す。

```
<?xml encoding='UTF-8'?>
<?xml encoding='EUC-JP'?>
```

SMS XMLプロセサによる実体及び参照の扱い

次の表に、文字参照、実体参照及び解析対象外実体の呼出しが現れる文脈、並びに、それぞれの場合における`wlk`プロセサに要求される振舞いを要約する。一番左の列のラベルは、参照が現れる文脈を示す。

内容における参照

要素の**開始タグ**及び**終了タグ**の間の任意の場所での参照。非終端記号`content`に対応する。

属性値における参照

開始タグの属性の値、又は**属性宣言**におけるデフォルト値のいずれかでの参照。非終端記号`Attvalue`に対応する。

属性値として出現

参照ではなく、`mame`として出現。`ENTITY`型として宣言した属性の値として出現するか、又は`ENTITIES`型として宣言した属性の値におけるスペースで区切るトークンの一つとして出現する。

実体値における参照

実体の宣言における、パラメタ実体又は内部実体の**リテラル実体値**の中での参照。非終端記号`dntityvalue`に対応する。

cscにおける参照

`csc`の内部サブセット又は外部サブセットでの参照。ただし、`dntityvalue`又は`Attvalue`の外側とする。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

SMSISIII “禁止”

次は禁止されており、**致命的エラー**とする。

aH 解析対象外実体への参照の出現。

bH cscのdntityualue又はAttualue以外の部分における、文字参照又は一般実体への参照の出現。

cH 属性値内の外部実体への参照の出現。

SMSITMI リテラル内での取込み

実体参照が属性値の中で現れたとき、または、パラメタ実体への参照がリテラル実体値の中で現れたとき、置換テキストは、参照自体の代わりに、参照があった位置に文書の一部としてあったものとして処理される。ただし、置換テキストの中の一重引用符又は二重引用符文字は、常に通常の文字データとして扱われ、リテラルを終了させることはない。例えば、次の文書例は整形形式である。

```
<!ENTITY YN "Yes" >
<!ENTITY WhatHeSaid "He said &YN;" >
```

z訳注Gこれは原文にはない) | 一行目は原規定ではパラメタ実体となっているが明らかな誤りなので修正した。

一方、次の例は整形形式ではない。

```
<!ENTITY EndAttr "27" >
<element attribute='a-&EndAttr;'>
```

SMSIGII “通知”

解析対象外実体の名前が、ENTITY型又はENTITIES型の属性値においてトークンとして現れたとき、妥当性を検証するプロセサは、アプリケーションに対して、その実体及び関連する記法のシステム識別子並びにG存在すればH公開識別子を通知しなければならない。

SMSI7II “処理しない”

一般実体参照が、実体宣言におけるdntityualue内に現れるとき、一般実体参照は処理されないで、そのまま残る。

SMSIMVII “PEとして取込み”

外部解析対象実体の場合と同様に、パラメタ実体は、妥当性を検証するときだけ取り込む必要がある。パラメタ実体参照をcsc内に認識して取り込むとき、その置換テキストは、その前後に一つのスペース文字GBxQOHの付加によって引き伸ばされる。パラメタ実体の置換テキストがcsc内の文法的トークンを完全に含むようにすることを、この規程は意図している。

SMTMI 内部実体置換テキストの構築

内部実体の取扱いの規定で、実体値を二つの形式に区別することは役に立つ。リテラル実体値は、実体宣言内に実際に存在する、引用符で囲まれた文字列とする。これは、非終端記号dntityualueとマッチする。置換テキストは、文字参照及びパラメタ実体参照の置換え後における、実体の内容とする。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

記法宣言は、記法の名前及び外部識別子を提供する。この名前は、外部実体宣言、属性リスト宣言、及び属性指定に用いる。外部識別子は、与えられた記法のデータを処理できるソフトウェア（ヘルパアプリケーションなど）を、w i kプロセサ又はクライアントアプリケーションが探すために利用できる。

```
zWQ|      motationcecl  Y= f[@mnsAshnmf r name r Gdxte-nalhc |oublichH r^f]f
```

```
zWR|      oublichc  Y= fot akhbf r oubidkite-al
```

w i kプロセサは、宣言されていて、属性値、属性定義又は実体宣言で参照されているすべての記法について、w i kプロセサは、記法の名前及び外部識別子をアプリケーションに提供しなければならない。さらに、外部識別子を、**システム識別子**、ファイル名又はその他の情報に展開してもよく、これらを用いて、アプリケーションは、その記法のデータを処理するプロセサを起動する。しかし、w i kプロセサ又はアプリケーションが動作するシステムでは利用できない記法を、w i k文書が宣言し参照しても、これは、エラーとはしない。

SMWM 文書実体

文書実体は、実体の成す木構造のルートであって、w i kプロセサが処理を開始する対象とする。この標準情報Gs qHは、w i kプロセサが、文書実体の存在する場所をどのように見つけるかは規定しない。他の実体と異なり、文書実体は名前をもたず、いかなる識別もなしにプロセサへの入力ストリームに出現してもよい。

TM 適合性

TMPM 妥当性を検証するプロセサ及び検証しないプロセサ

適合w i kプロセサは、妥当性を検証するもの及び妥当性を検証しないものの二つに分類される。

妥当性を検証するプロセサも妥当性を検証しないプロセサも、読み込んだ文書実体及び他のすべての解析対象実体において、この標準情報Gs qHの整形形式制約への違反を報告しなければならない。

妥当性を検証するプロセサは、c s c内の宣言によって示された制約への違反と、この標準情報Gs qHが規定する妥当性制約への違反とを、すべて報告しなければならない。これを実現するために、妥当性を検証するw i kプロセサは、c s c全体と文書内で参照されているすべての外部解析対象実体とを読み込んで処理しなければならない。

妥当性を検証しないプロセサは、整形形式であることを確認するために、c s cの内部サブセット全体を含めた**文書実体**を調べることだけが義務づけられている。文書の妥当性を確認する必要はないが、読み込んでいないパラメタ実体への参照が最初に起きるまでに読み込んだc s cの内部サブセットとパラメタ実体とに現れるすべての宣言を処理しなければならない。すなわち、属性値を**正規化**し、内部実体の置換テキストを**取込み**、**デフォルトの属性値**を与えるために、これらの宣言にある情報を使用しなければならない。実体の宣言は上書きされる可能性があるため、妥当性を検証しないプロセサは、読み込んでいないパラメタ実体への参照より後に現れた**実体宣言**及び**属性リスト宣言**を**処理**してはならない。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

"string"

二重引用符で囲むリテラル文字列とマッチするリテラル文字列とマッチする。

'string'

一重引用符で囲むリテラル文字列とマッチするリテラル文字列とマッチする。

これらの記号は、次の形式の組合せで使用する。ここで、A及びBは式とする。

GexpressionH

ここに示す組合せによる式GexpressionHを、一つのまとまりとして扱うために使う。

A?

Aのオプションな出現とマッチするGオプションのAH。

A B

Aの次にBが出現するものとマッチする。

A | B

A又はBのどちらかとマッチする。

A - B

AとマッチするがBとはマッチしない任意の文字列とマッチする。

A+

AのP回以上の繰返しとマッチする。

A*

Aの0回以上の繰返しとマッチする。

生成規則内で使用する他の記法を次に示す。

/* ... */

コメント。

[wfc: ...]

整形形式制約。生成規則に関連した、[整形形式](#)の文書に関する制約を名前によって特定する。

[vc: ...]

妥当性制約。生成規則に関連した、[妥当な文書](#)に関する制約を名前によって特定する。

附属書 A参考文献

A.1 規定の参考文献

IANA

Ghnte-net Assigned numbe-s Autho-ityH nfficial mames fo- bha-acte- retsK edll j eld rimonsen et alll <ftp://ftp.isi.edu/inl/notes/iana/assignments/cha-acte-lsets>を参照。

IETF RFC 1766

hdse Ghnte-net dnginee-ing sask eo-ceHll qeb PVUUY sags fo- the hdentification of kanguagesK edll gll Alvest-andll PXXTll

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

IETF RFC2141

hdse Ghnte-net dnginee-ing sask eo-ceH|| qeb QPSY t qm ryntaxK ed|| q|| I oats|| PXXV||

ISO 8879

hr n Ghnte-national n-ganization fo- rtanda-dizationH|| hr n WWWXLPXWU GdH|| hnfo-mation p-ocessing LL sext and nffice systems LL rtanda-d f ene-alized I a-kup kanguage Gr f I kH|| ei-st edition LL PXWULPOLPT|| z f eneva|Y hnte-national n-ganization fo- rtanda-dizationK PXWU||

ISO/IEC 10744

hr n Ghnte-national n-ganization fo- rtanda-dizationH|| hr nNhd b POVSSLPXXQ GdH|| hnfo-mation technology LL gype-mediaNsimeLbased rt-uctu-ing kanguage GgysimeH|| z f eneva|Y hnte-national n-ganization fo- rtanda-dizationK PXXQ|| dxtended eacilities Annexel|| z f eneva|Y hnte-national n-ganization fo- rtanda-dizationK PXXU||

附属書 B文字クラス

t nicode標準に定義するプロパティにしたがって、文字は基底文字Gaasebha-H|これらは、発音符を除くラテンアルファベットのアルファベット文字を含む。H、統合漢字Gideog-aphicH及び結合文字Gbombiningbha-H|このクラスはほとんどの発音符を含む。Hにクラス分けする。これらのクラスを合わせて字Gkette-Hのクラスとする。PO進数値GcigithH及びエクステンダGdxtende-Hのクラスもある。

zWS	kette- Y=	aasebha- hdeog-aphic
zWT	aasebha- Y=	zBx00SPLBx00TA zBx00UPLBx00VA zBx00b0LBx00cU zBx00cWLBx00eU zBx00eWLBx00ee zBx0POOLBx0PRP zBx0PRSLBx0PRd zBx0PSPLBx0PSW zBx0PSALBx0Pvd zBx0PWOLBx0PbR zBx0Pb cLBx0PeQ zBx0PeSLBx0PeT zBx0PeALBx0QPv zBx0QTOLBx0QAW zBx0QaaLBx0QbP BxORWJ zBxORWWLBxORWA BxORWb zBxORwdLBxORAP zBxORARLBxORbd zBxORc0LBxORcU BxORcA BxORc b BxORc d BxORd d zBxORdQLBxOReR zBxOSOPLBxOSOb zBxOSOdLBxOSSe zBxOSTPLBxOSTb zBxOSTdLBxOSWP zBxOSXOLBxOSb\$ zBxOSbVLBxOSbW zBxOSbaLBxOSbb zBxOSc0LBxOSda zBxOSddLBx0SeT zBx0SeWLBx0SeX zBxOTRPLBx0TTU Bx0TTX zBx0TUPLBx0TWU zBx0Tc0L Bx0TdA zBx0Te0LBx0TeQ zBx0UQPLBx0OURA zBx0USPL Bx0USA zBx0UVPLBx0UaV zBx0UaALBx0Uad zBx0Ub0L Bx0Ubd zBx0Uc0LBx0UcR Bx0UcT zBx0UdTLBx0UdU zBx0XOTLBx0XRX Bx0XR c zBx0XTWLBx0XUR zBx0XWTL Bx0XWb zBx0XWeLBx0XXQ zBx0XXRLBx0XAW zBx0XAAL Bx0Xa0 Bx0XaQ zBx0XaULBx0XaX zBx0XcbLBx0Xcc zBx0XcelBx0XdR zBx0Xe0LBx0XeR zBx0AOTLBx0A0A zBx0A0eLBx0APQ zBx0APRLBx0AQW zBx0AQALBx0ARO zBx0ARQLBx0ARR zBx0ARTLBx0ARU zBx0ARWLbX0ARX zBx0ATXLBx0ATb Bx0AT d zBx0AVQLBx0AVS zBx0AWTLBx0AWa Bx0AW c zBx0AWeLBx0AXP zBx0AXRLBx0AAW zBx0AAALBx0AaQ zBx0AaQL Bx0AaR zBx0AaTLBx0AaX Bx0Aa c Bx0Ad0 zBx0a0TLBx0a0b zBx0a0eLBx0a0PQ zBx0aPRLBx0a0QW zBx0aQALBx0a0RQ zBx0aRQLBx0a0RR zBx0aRULBx0a0RX

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

```

| zBx0badLBx0bbS| zBx0bbULBx0bbW| zBx0bbAL
Bx0bb c|| zBx0b cTLBx0b cU| zBx0cOQLBx0cOR|
| zBx0cRdLBx0cSR| zBx0cSULBx0cSW| zBx0cSALBx0cSc|
| Bx0cTV Bx0dRP zBx0dRSLBx0dRA| zBx0dSVLBx0dSd|
| Bx0daP zBx0daSLBx0daX| zBx0daaLBx0dab|
| zBx0dbWLBx0db c| zBx0ePWLbX0ePX| Bx0eRT Bx0eRV
| Bx0eRk Bx0eRd Bx0eRe zBx0eVPLBx0eWS| zBx0eWUL
Bx0eWa|| zBx0eXOLBx0eXT| Bx0eXV zBx0eXXLBx0eAc|
| zBx0eaPLBx0eaV| Bx0eaX zBx0Q0c0LBx0Q0c b| Bx0Q0P
| zBxROQALBxROQe| BxROXk BxROXA
zWW|          cigit Y= zBx00R0LBx00RX|| zBx0UU0LBx0UUx| zBx0Ue0LBx0UeX|
| zBx0XUULBx0XUe| zBx0XdULBx0Xde| zBx0AUULBx0AUe|
| zBx0AdULBx0Ade| zBx0aUULBx0aUe| zBx0adVLBx0ade|
| zBx0bUULBx0bUe| zBx0bdULBx0bde| zBx0cUULBx0cUe|
| zBx0dTOLBx0dT| zBx0dc0LBx0dcX| zBx0eQ0LBx0eQX|
zWX|          dxtende- Y= Bx00aV| Bx0QcD Bx0QcP BxORWV Bx0USP Bx0dSU
| Bx0dbU BxROO| zBxRORPLBxRORT| zBxROXcLBxROXd|
| zBxROebLBxROed|

```

ここで定義する文字クラスは、t nicode文字データベースから、次のとおりを得ることができる。

aH 名前開始文字は、k k ku k ko k kt k mlカテゴリ内の一つでなければならない。

bH 名前開始文字以外の名前文字は、l c k l e k l n k km k mdカテゴリ内の一つでなければならない。

cH 互換性領域にある文字G文字符号でBxeX00より大きくBxeeedより小さい文字Hは、w l kにおける名前としては許されない。

dH フォント分解か互換性分解をもつ文字Gつまり、データベース内の5番目のフィールドにAcompatibility fo-matting tagAがあるもの。これは、5番目のフィールドが、A[Aで始まることによって示される。Hは許されない。

eH 次の文字は、名前開始文字として扱う。これは、プロパティファイルが、これらの文字をアルファベットに類似すると見なすことによる。それらはzBx0QaaLBx0QbP|k Bx0TTXk Bx0UdTK Bx0UdUとする。

fH 文字符号がzBx0Q0c cLBx0Q0d0|の文字は、G t nicode のTMPSに従ってH除外する。

gH 文字符号がBx00aVの文字は、プロパティリストにしたがって、エクステンダGextende-Hに分類する。

hH 文字BxORWVは、これに相当する標準形がBx00aVなので、名前文字に追加する。

iH 文字YF及びFは、名前開始文字として使ってよい。

jH 文字Rk及びRkFは、名前文字として使ってよい。

附属書 CXML及びSGML【非準】

w l kは、r f l kのサブセットとして設計されている。すなわち、すべての妥当なw l k文書は、規格に適合するr f l k文書にもなる。r f l kが文書に課す制限以外に、w l kがいかなる制限を課すかについての詳細は、zbla-k|を参照のこと。

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

dHW行目で、一般実体AtrickyAへの参照を認識し、展開する。要素AtestAの完全な内容は、次の自己記述的なG非文法的なH文字列となる。つまり、shis sample sho•sa e—o—lp—one methodll

附属書 E 決定的内容モデル【非準】

互換性のため、要素型宣言における内容モデルは、決定的でなければならない。

r f l kは、決定的内容モデルr f l kでは、非あいまいと呼ぶ。Hを要求する。r f l kシステムを用いて作成したw l kプロセサは、非決定的内容モデルをエラーとしてもよい。

例えば、内容モデル((b, c) | (b, d))は非決定的となる。これは、最初にbを与えたとき、モデル内のいずれのbとマッチするのか、その次の要素を先読みすることなしには、パーサは知ることができないことによる。この場合は、bの二つの出現は、一つにまとめることができ、モデルは(b, (c | d))となる。こうすれば、明らかに最初のbは、内容モデル内の一つの名前とだけマッチする。パーサは先読みして、次にくるものを知る必要がない。cもdも受理される。

形式的に示せば次の通り。AhoK rethik and t llman zAhoK t llman|のR1X節のアルゴリズムR1Tなどの標準的なアルゴリズムを用いて、内容モデルから有限オートマトンを構成することができる。この種の多くのアルゴリズムでは、正規表現における各々の位置Gつまり、正規表現の構文木における各々の末端ノードHに対して、follo• setを構成する。ある位置に対するfollo• setにおいて、複数の位置が同じ要素型名でラベル付けされていれば、その内容モデルはエラーとなり、エラーとして報告されることもある。

すべての非決定的内容モデルを等価な決定的内容モデルに変換することはできないが、多くの非決定的内容モデルを変換するアルゴリズムが存在する。a-üggemannL j lein PXXP za-üggemannL j lein|を参照のこと。

附属書 F 文字符号化方式の自動検出【非準】

w l kの符号化宣言は、各実体の内部ラベルとして機能し、どの文字符号化方式を使用するかを示す。しかし、w l kプロセサは内部ラベルを読む前にどの文字符号化方式が使われているかを知る必要があり、これが、内部ラベルが示そうとしていることに他ならない。一般的には、これは絶望的な状態となる。しかし、w l kにおいては、完全には絶望的ではない。これは、w l kが次の二つの点で一般的な場合に対する制限を加えていることによる。一つの制限は、どの実装も有限個の文字符号化方式だけをサポートするものと見なす。他の一つは、w l kの符号化宣言の位置及び内容を制限して、各実体で使用する文字符号化方式の自動検出を可能にする。また、多くの場合に、w l kのデータストリームに加え、他の情報が利用できる。ここでは、w l kの実体がプロセサに渡される時、G外部H情報を伴うかどうかによって、二つの場合に分ける。まず最初の場合を示す。

t selW形式又はt selPU形式ではないw l k実体は、最初の文字列を‘ <?xmlfとするw l k符号化宣言で始めなければならないので、どの適合したプロセサも、入力にあるQオクテット又はSオクテットを調べれば、次のどの場合があてはまるかを検出できる。このリストを読む際には、t brLSのFfがA#x0000003CA、FfがA#x0000003FA、及びt selPUのデータストリームの必要とするバイト順マークがA#xFEFFAということを知っておくと役立つ。

aH 00 00 00 3CY t brLSK bigLendian マシン GPQRS順H

bH 3C 00 00 00Y t brLSK littleLendian マシン GSRQP順H

XEP Trial Version limitation:

starting from page 11, all odd pages are blank.

るのがよい。例えば、内部ラベル及び外部ヘッダに存在する l h l d形式のラベルの相対的な優先度についての規則は、l h l d型textlxml及びapplicationlxmlを定義するqeb文書の一部となるのが望ましい。しかし、相互運用性のため、次の規則を推薦する。

a h w l kの実体がファイルに存在すれば、バイト順マーク及び符号化宣言ohは、G存在すればH文字符号化方式を決定するために使用する。他のすべてのヒューリスティック及び情報は、エラー回復のためだけに用いる。

b h w l kの実体を l h l d型textlxmlで配送するときは、この l h l d型のもつcha-setパラメタが文字符号化方式を決定する。他のすべてのヒューリスティック及び情報は、エラー回復のためだけに用いる。

c h w l kの実体を l h l d型applicationlxmlで配送するときは、バイト順マーク及び符号化宣言ohをG存在すればH文字符号化方式の決定のために使用する。他のすべてのヒューリスティック及び情報はエラー回復のためだけに用いる。

これらの規則は、プロトコルについての資料がないときにだけ用いる。特に、l h l d型textlxml及びapplicationlxmlが規定されれば、関連qebの勧告が、これらの規則に取って代わる。

附属書 GW3C XML作業グループ【非準】

この標準情報Gs qHの原勧告は、v R b w l k作業グループG v f Hが準備し、公開を承認した。v fがこの標準情報Gs qHの原勧告を承認するということは、v fのすべての委員が承認投票を行ったということを必ずしも意味しない。w l k v fの現在の委員及び以前の委員を次に示す。

ion aosakK run Gbhai-HZ iames bla-k Gsechnical keadHZ sim a-ayK sexuality and metscape Gw l k bol edito-HZ iean oaolik l ic-rosoft Gw l k boLedito-HZ bll l rpe-be-gL l cpueenK tll of hllll Gw l k boLedito-HZ can bonnollyK v R b G v R b kaisonHZ oaula Ange-steinK sexcelZ rteve ceqoseK hm r nZ cave gollande-K goZ dliot jimbe-K hr n f dmZ dve l ale-K A-bo-sextZ som l aglie-yK mbrAZ l u-ay l aloneyK l uzmo and f-ifz 村田 真, 富士ゼロックス情報システムG株HZ ioel mavaK AdobeZ bonleth nfbonnellK uignetteZ oete- rha-peK roftpuadZ iohn sigueK catabhannel