2007－CE－88 (4)
2007／2／16

XMLを用いた技術史の知識基盤表現

松井 正志　中平 勝子　三上 喜貴
長岡技術科学大学

抄録

従来の歴史記述は人間が読むことを前提として記述されているため，コンピュータが自動的に収集・加工できるような記述が行われていない．コンピュータが歴史イベントの記述を認識し，自動的に処理できるようになることで，目的にあった歴史教材の作成が可能になる．本稿では，フィルモアの格文法における深層格をベースに既存の歴史記述言語の比較評価を行い，歴史的イベントの記述枠組みを提案した．また，これを元にmicroformatsの語彙を作成した．

Knowledge Representation of Technology History Using XML

MATSUI Masashi　NAKAHIRA T. Katsuko　MIKAMI Yoshiki
Nagaoka University of Technology

Abstract

Traditional contents on historical knowledge are written for human readers, and are not written in a way which enables computers to handle it. If computers can "read" historical events and process it automatically, teaching materials can be easily compiled depending on requirements. The paper reviews a few existing description frameworks for historical events, and makes a comparative assessment of the usability of those frameworks based on "deep core" of Fillmore's case grammar. A new description framework is proposed based on this study, and a set of microformats' vocabularies is created.

1. はじめに

現在のWeb上のコンテンツは，人間にとって読みやすいものであるが，コンピュータにとって読みやすいものではない．そこで，Web上のコンテンツにコンピュータが理解可能な意味を持たせる，セマンティックWebというプロジェクトがW3Cにより進められている．セマンティックWebのゴールは，Webを人間が読むための「文書のWeb」からコンピュータも理解可能な「データのWeb」にすることである[1]．

「データのWeb」を教育に利用しようと考えた時，その有効利用法の一つに「資料の自動収集・整理・表現」が考えられる．例えば歴史について考えると，文献，年表やデータベースなど様々なデータがWeb上にあるが，コンピュータが意味を解釈し，自動的に収集を行える記述はされていない．歴史的イベントの情報を自動的に収集および再利用可能にすることで，歴史の多角的な視点からの分析に発するような教材作成が可能となる．

編集工学研究所の開発した「クロノス・システム」[2]は，同じ時間にある複数分野の歴史を同時に見ることができ，三次元空間の中に歴史的イベント配置することで，時代の流れと事項間の関連などの広がりを見ることができる．また，クロノス・システムは，自分で歴史のデータを作ることもできる．そして作った歴史を線で繋げ，歴史間の関係を作ることができる．しかし，クロノス・システムの場合は，クロノス・システム上で作られたデータしか扱うことができない． Web上にある豊富な情報資源を利用し，クロノス・システムのような多角的な視点での分析が行えるシステムを実現するには，土台として歴史的イベントを記述する枠組みを決める必要がある．本稿では，特に因果関係をはっきりつけやすい技術史に対してその知識基盤表現を試みる．

2. これまでの研究

これまでに，国内外を問わず歴史的知識を記述する形式や言語を開発する試みが行われてきた．国内では，電気学会の電気技術史技術委員会が 1996 年に電力系統技術歴史データベースの作成[3]を行い，そのための歴史記述言語HSML (Historical Space Modeling Language)と，HSMLで記述された歴史情報を閲覧するGUI「曼荼羅」を開発した[4]．海外ではHEML (Historical Event Markup and Linking)プロジェクト[5]が，歴史記述を目的とした独自のXMLスキーマと，それを年表や地図など様々な形式で表示するWebアプリケーションを開発している．また，歴史記述に近い概念として，スケジュール管理の分野ではイベントの記述形式にiCalendar[6]が用いられてきた． Webベースのソフトウェアが普及してきた現在，

図 1. Knecht クロニクル

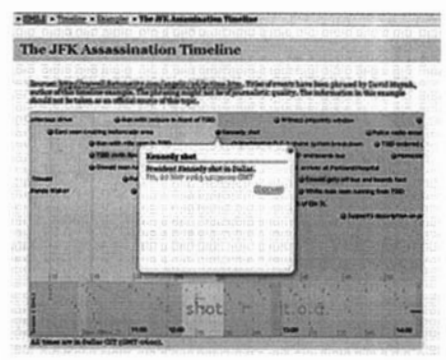

図 2. SIMILE Timeline

iCalendarをRDFで記述するRDF Calendarや，XHTML中にiCalendarと同等の項目を埋め込むmicroformats「hCalendar」など，iCalendarをWeb技術と親和性の高い記述方法で記述する試みが行われている．また，その他にも書誌情報のメタデータ記述の枠組みとしてDublin Core[7]があるが，歴史の分野でも国立歴史民族博物館の歴史研究データベースの項目をDublin Coreの語彙にマッピングする試み[8]が行われた．このようにWeb上で歴史やイベントを記述する様々な試みが行われている．

記述方式が開発される一方で，歴史的知識を直感的にわかりやすい形に表現する様々な試みが行われている．国内では

KnechtがGoogleやWikipediaなどのWebサービスを組み合わせ，歴史的出来事を地図上に表示するサービス「Knechtクロニクル」[9]を開発した．海外ではW3C，MITコンピュータ科学・人工知能研究所，MIT図書館の３機関による共同プロジェクトSIMILEがAJAXを用いて，ドラッグによる直感的な操作が可能な年表表示ツール「Timeline」[10]を開発した．

このように，記述言語開発と視覚化の両面で歴史的イベントを知識基盤として表現する試みが行われている．しかし，因果関係を扱えるものは少ない．前述した中では電気学会の HSML・曼荼羅のみである．しかしHSMLは，HSMLインタプリタがないと利用できないため，現在広く普及しているRDFなどの記述形式と比較すると，汎用性の面で難がある．よって，本稿では汎用性のある記述形式の開発を試みる．

3. 記述項目の検討

本稿では，記述項目の検討として，文の意味解析で使われるフィルモアの深層格を用いた．また，既存の記述形式について，深層格から導き出された項目にどの程度対応しているか比較した．

3.1. 格文法の深層格

歴史的イベントの記述項目の検討枠組みとして，本稿ではフィルモアの格文法における深層格をとりあげた．深層格は，動詞に対して個々の単語がどのような役割を持つかを示す．歴史的イベントには，少なくとも何らかの動作と動作に対応づけることが可能である．したがって，深層格の与える枠組みはイベント記述に関する最大限の記述項目を提示すると考えられる．フィルモアの深層格の与える8つの格を表1に示す．

8つの格を歴史的イベント記述にあてはめた場合，それぞれの格は次のように対応する．動作主格はイベントを発生させた主体に当たる．それは個人であることもあれば，組織であることも考えられる．道具格はそのイベントが起きた原因に当たる．対象格はイベントの対象だが，これはイベントの内容による．例えば，新技術の開発などの場合は，その技術が対象格に該当する．場所格はそのイベントが起きた場所，時間格はそのイベントが起きた時間に当たる．イベントはいわば直接話法によって表記されるため，経験者格は存在しない．また，源泉格と目標格の解釈は難しい．時間や場所と違い，始点と終点があるかはイベント

表1. フィルモアの深層格

格	説明
動作主格	ある動作を引き起こす者の役割.
経験者格	ある心理事象を体験する者の役割.
道具格	ある出来事の直接原因となったり，ある心理現象と関係して反応を起こさせる刺激となる役割
対象格	移動する対象物や変化する対象物．あるいは，判断，想像のような心理事象の内容を表す役割.
源泉格	対象物の移動における起点，および状態変化と形状変化における最初の状態や形状を表す役割.
目標格	対象物の移動における終点，および状態変化と形状変化における最終的な状態や結果を表す役割.
場所格	ある出来事が起こる場所および位置を表す役割.
時間格	ある出来事が起こる時間を表す役割.

の性質（動詞）に特に依存している．そのため本稿では，イベントにおいて源泉格と目標格に該当するものはないものとした．

これらの結果から，本稿では歴史的イベントを構成する要素を「人物・原因・対象・場所・時間」とした．

3.2. 既存の記述形式の比較

イベント記述項目の検討材料としてiCalendar，HEML，Timeline，Dublin Coreの比較を行った．比較を行った結果が表2である．

個々の項目に記述できる内容に着目すると，地理情報の緯度経度や関与者の役職の項目など，HEMLはより詳細な記述方法が用意されている．Timelineは他の記述形式に比べて簡潔なものとなっているが，これはTimelineは年表表示という機能に特化しているからである．Timelineの場合，3.1.で挙げた要素は全て説明の中に記述される．

また，どの記述形式も参照情報へのリンクを記述できる．説明は，深層格のどれにも当てはまらない．説明はイベントそのものを表しているため，深層格の全てを含む．よって，説明は構造化した歴史イベントの項目として含むべきではない．歴史的イベントに関する知識という膨大な情報の記述は階層構造を持って作られる必要があり，詳細については下位の階層（参照情報の示す参照先）へ記述するのが現実的である．

Dublin Core以外の記述形式は対象を記述する項目が無かったが，多くの場合イベント名は「対象＋動詞」という形をとるため，対象に当たる項目はイベント名に含まれていると考える．

表2を見るとDublin Coreが最も記述項目が多いが，元々書誌情報に対して設計されているものをイベントに当てはめたので無理が生じる部分がある．例えば，本稿ではcreator（作者）の項目に人物を記述できると解釈した．しかし，例えば歴史的イベントの一つとして「関が原の戦い」について考えたとき，「関が原の戦い」の作者とは誰だろうか．また，イベントの時間は有効期日（date）ともの時間的対象（coverage）と

表2. 既存のイベント記述形式の比較

深層格	項目名	HEML	iCalendar	Dublin Core	Timeline
動作主格	人物	名前，役職を記述可	イベントの主催の名前や連絡先を記述可	記述可	－
道具格	原因	－	－	－	－
時間格	時間	開始と終了を記述可	開始と終了を記述可	イベント生成時間と範囲を記述	開始と終了を記述可
場所格	場所	地名，緯度経度で記述可	地名を記述可	地名，緯度経度で記述可	－
対象格	対象	－	－	記述可	－
－	イベント名	記述可	記述可	記述可	記述可
－	説明	－	記述可	記述可	記述可
－	参照情報	記述可	記述可	記述可	記述可
－	その他	多言語対応可	優先度や関連イベントの記述が可	書誌情報をベースにしているため，権利者に関する項目などがある	文字色やアイコンなど，表示に関する記述

も解釈できるがどちらが適切だろうか．リソースの種類を示すtype要素のための語彙として，Dublin CoreはDCTYPEという12の要素を定義している．その中にはEventがあるが，実際のイベントへの適用を考えると様々な解釈上での問題があり，時には拡張を行わなくてはならない．そのためイベントにDublin Coreを適用する場合，十分な検討を行う必要がある．

イベントに関する記述項目について，どの記述形式も基本的な事項を書くには十分な表現力を持っている．しかし，原因について記述を行える記述形式はない．本稿では，歴史記述に因果関係を含める．そのため，原因を記述できる記述形式を開発する必要がある．

3.3. 記述項目

3.1.で述べた要素や，3.2.で行った比較を元に記述項目を作成したのが表3である．対象については，イベント名に「対象＋動詞」と記述されることを想定し，項目としなかった．

4. 記述方法の検討

本稿では歴史イベントのデータ形式として現在広くWebで利用されているRDFを用いた．しかし，RDFはコンピュータにとって読みやすい反面，人間にとっては読みづらい．そのため本稿では，HTML文書中の属性などにマークアップを行い，HTML文書からコンピュータが読みやすいデータを抽出しRDF形式で出力する，という形を取った．

表3. 本稿の提案する記述項目

分類	項目	データ型
イベント名	イベント名	文字列型
時間	開始期日	日付型
	終了期日	日付型
場所	地名	文字列型
	緯度	数値型
	経度	数値型
人物	人物名	文字列型
	人物URL	URL
原因	原因名	文字列型
	原因URL	URL
参照情報	参照情報名	文字列型
	参照情報URL	URL

4.1. microformats

本稿は，コンテンツ内のメタデータの記述方法にmicroformatsを用いた．microformatsは，XHTMLの属性値に一定の名前づけを行うことで，文書からメタデータの抽出を行うものである．microformatsはXHTMLの文法に用意されている属性値を利用する．そのため，現存するコンテンツに対して適用してもその構造をほとんど変更せずに済むという利点がある．本稿では，従来ある歴史的イベントの記述に対しての適用も考慮しているため，このような利点を持つmicroformatsによってイベント記述を行った．

microformatsの仕様には，iCalendarをmicroformatsとして実装したhCalendarがある．しかし，hCalendarでは表3に示した緯度経度，原因などの項目を記述できない．そのため，本稿では新たにmicroformatsの語彙を定義した．

また，microformatsの語彙を定義すると共に，定義したmicroformatsのメタデータを抽出し，RDFに変換するXSLTを作成した．XSLTはXML文書を他のXML文書に変換する変換言語である．本稿ではXSLTをXHTMLのlink要素中で指定することで，コンテンツにXSLTファイルを関連付けている．エージェントはこのXSLTファイルを用いて，自動的にメタデータを抽出できる．

このような文書にコンピュータがメタデータを抽出するためのXSLTファイルを関連付ける仕組みをGRDDL (Gleaning Resource Descriptions from Dialects of Languages)と呼ぶ．W3CによりGRDDLワーキンググループが組織される[12]など，今後GRDDLに関する様々な仕様・技術・活用例などが生み出されることが期待される．

4.2. 語彙の作成

3.で検討した結果から，表 4 のようなmicroformats の語彙を定義した．

イベントは，一つの要素で表され，その要素は class 属性に「event」の値を持つ．また，イベント名はその要素の title 属性内に記述される．以降のイベントに関する情報は，全てこの要素内に記述される．

イベントが起きた時間は，class 属性に「start_date」の値を持つ要素の内容と，title 属性の値によって記述される．要素の内容は「19 世紀初頭」や「2004 年 10 月 23 日、午後 5 時 56 分」などの値が入り，数値・文字列は問わない．一方 title 属性の内容は日付型で記述する．具体的には，XML Schema で定められている dateTime 型，date 型，gYearMonth 型，gYear 型のいずれかで記述する．また，イベントに時間的範囲がある場合は，class 属性に「end_date」の値を持つ要素で，範囲の終了を示す．内容の記述方法については，イベント開始時間と同様である．

イベントが起きた場所は，class 属性に「location」の値を持つ要素の内容と，その要素の title 属性によって示す．要素の内容は地名を示す．title 属性に緯度経度を記述するが，記述方法としては「緯度，経度」の順にカンマ区切りで記述する．

人物，証拠，原因は，a 要素を用いる．rel 属性にそれぞれ定義した値を入れ，要素の title 属性に名前，href 属性に URL を記述する．関与者，証拠については対象となる URL がない場合があるが，その場合は任意の要素の class 属性に rel 属性と同じ値を入れ代替するものとする．証拠については，書籍など URL で表せないものが多数あると考えられるが，これについては今後検討するものとする．

5. 試作

試作として，技術史の文献を 4.2.で示し

表 4. 本稿の提案する microformats の語彙

分類	要素	属性	値	項目	記述箇所	データ型	備考
イベント	指定なし	class	event	イベント名	title属性	文字列型	
時間	指定なし	class	start_date	開始期日名	要素内	文字列型	
				開始期日	title属性	日付型	
	指定なし	class	end_date	終了期日名	要素内	文字列型	期日に範囲の無い場合は省略.
				終了期日	title属性	日付型	
場所	自由	class	location	地名	要素内	文字列型	
				緯度経度	title属性	数値型	緯度経度をカンマ区切りで記述.
関与者	a	rel	participant	関与者名	title属性	文字列型	
				関与者URL	href属性	URL	
証拠	a	rel	evidence	証拠名	title属性	文字列型	
				証拠URL	href属性	URL	
原因	a	rel	cause	原因名	title属性	文字列型	
				原因URL	href属性	URL	

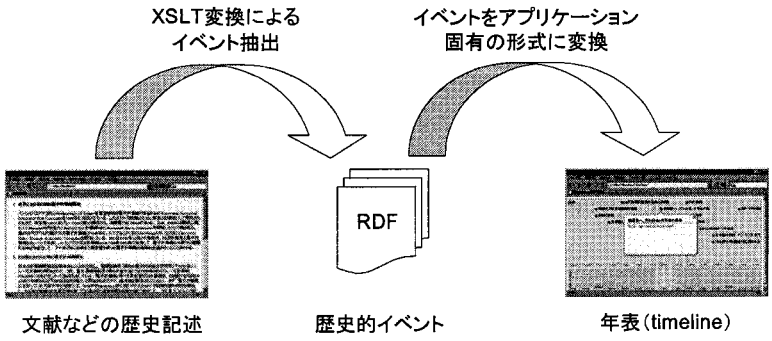

図 3. 試作の処理の流れ

た microformats の語彙でマークアップし，マークアップされたイベントを SIMILE の timeline 上で表示した．本稿では試作の対象として，電気学会の報告書の「世界における初期の電子計算機開発」と「日本における初期の電子計算機の開発」を用いた．

試作の大まかな流れを示したのが図 3 である．まず，定義した語彙を用いて文献をマークアップする．そして，文書に関連付けられている XSLT を用いて文書からイベントを抽出し，イベントが記述された RDF を取得する．そして，表示を行うアプリケーションを年表上にマッピングできることを確認した．

Timeline のデータは独自の語彙を用いた XML 形式である．よって，microformats から直接 Timeline 形式に変換する XSLT を作成すれば，文書から Timeline のデータを生成することも可能である．しかし，本稿では RDF に一旦変換してから，Timeline 形式に変換している．これは特定の表示形式への対応ではなく，様々な表示形式での利用を想定しているからである．

試作の結果，図 4 のように抽出したイベ

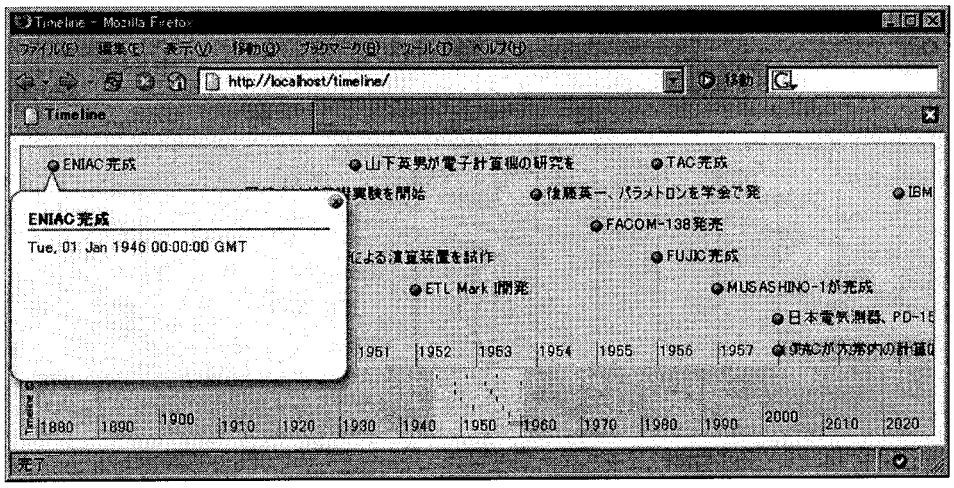

図 4. 試作の結果

6. まとめ

本稿は，歴史的イベントの記述枠組みを作成するために，フィルモアの格文法における深層格をベースに既存の歴史記述言語の比較評価を行い，これらに不足する項目を補った独自の記述枠組みを作成した．また，作成した枠組みに対応する microformats の語彙を作成した．この語彙を用いてマークアップされたイベントを直感的な形で表示するツールの開発を行うことが今後の課題である．

参考文献

[1] W3C Semantic Web Activity. [Online]. Available: http://www.w3.org/2001/sw/ (January 2007)

[2] クロノス・システム. [Online]. Available: http://www.eel.co.jp/02_core/002_cronos/cronos.html (January 2007)

[3] 電気技術史技術委員会，電気学会技術報告，第 991 号，pp.9 - 11

[4] Matsumoto, Y., A. Yamada, An association-based management of reusable software components, Annals of Software Engineering 5 (1998)

[5] Robertson, B.G, Visualizing An Historical Semantic Web with HEML, WWW2006, May 2006, Edinburgh.

[6] RFC 2445 - Internet Calendaring and Scheduling Core Object Specification (iCalendar) (1998) [Online]. Available: http://tools.ietf.org/html/rfc2445 (January 2007)

[7] Dublin Core Metadata Initiative. [Online]. Available: http://purl.oclc.org/dc/ (January 2007).

[8] 安達 文夫，歴史研究データベースの Dublin Core へのマッピングとその課題，人文科学とコンピュータ研究会報告 2006-CH-72 (2006)

[9] Knecht. Knecht クロニクル. [Online] Available: http://chronicle.knecht.jp/ (January 2007).

[10] SIMILE. Timeline. [Online] . Available: http://simile.mit.edu/timeline/ (January 2007).

[11] 長尾 真，自然言語処理，岩波書店 (2005)

[12] W3C GRDDL Working Group. [Online]. Available: http://www.w3.org/2001/sw/grddl-wg/ (January 2007).

[13] 神崎 正英，RDF/OWL 入門，森北出版 (2005)