

利便性の高い Semantic Wiki の構築とその評価

Implementation and Evaluation of a Convenient Semantic Wiki

児玉 英一郎†
Eiichiro Kodama

高橋 希世士‡
Kiyoshi Takahashi

王家宏†
Jiahong Wang

高田 豊雄†
Toyoo Takata

1. まえがき

インターネットの普及に伴い、Web 上には多数の情報が存在している。また最近では、これらの情報の中に、CMS (Contents Management System) によって作られた Web ページが数多く見受けられるようになった。CMS とは、Web ページなどの作成、管理を支援するシステムであり、代表的なものとして、Movable Type などの Weblog ツールや、Wiki[1]などが挙げられる。

Wiki とは、Cunningham によって考案された Wiki Wiki Web 自身、あるいは、これから派生した数多くの Wiki クローンを指しており、この Wiki を利用することで、Wiki ページと呼ばれる Web ページを容易に作成、管理することが可能となる。実際、Wiki では、Wiki 構文と呼ばれる独特な構文が利用でき(例えば、ある Wiki では、[[リンク先 Web リソースの URI]]といった簡易記法でリンクを記述可能)、HTML などの専門知識の習得負担を軽減している。

このように、Web は情報発信者にとって、容易に情報発信可能な状況にあり、Web リソースは増加の一途をたどっている。そして、その結果、情報受信者は必要な情報を効率的に発見することが困難になってきている。

このような状況を受け、Web 上のリソースを情報受信者に代わりソフトウェアによって自動的に処理するための枠組みとして、Semantic Web[2][3]という概念が注目されている。Semantic Web では、各 Web リソースに対し、RDF(Resource Description Framework)[4]に基づいたメタデータが付与される。そして、このメタデータを参照することで、ソフトウェアによる Web リソースの意味内容の解釈が可能になると考えられている。その結果、ソフトウェアによる Web からの効率的な情報の収集や高度な検索が可能になると期待されている。

このように Semantic Web は有用な可能性を秘めており、近年では、Semantic Web の技術を用いた Wiki に関する研究も行われている。このような Wiki は Semantic Wiki と呼ばれている。この Semantic Wiki は、通常、Wiki ページに対しメタデータの付与を支援するための機能を有しており、また、付与されたメタデータを利用した高度な検索機能を提供している。

しかし、既存の Semantic Wiki は、利用者に対し、従来の Wiki にはなかった煩雑な操作や、利用のための専門的な知識を要求するといった問題を持っている。また、検索を行った際に、検索結果中に多数の Wiki ページがあった場合、どの Wiki ページが適切であるか、利用者が内容を見て選定する必要がある、目的の情報を探しやすいとはいえない。

このため、本研究では、利用者がメタデータの生成や、メタデータを用いた検索を行う際に、従来の Wiki と同等の操作で専門的な知識の習得なしに利用でき、メタデータを用いた検索においては、目的情報の発見を容易とする

ような Semantic Wiki の構築を試みる。即ち、これを実現するための利便性の高い Semantic Wiki のモデルを提案する。本提案の Semantic Wiki のモデルに従い構築された Semantic Wiki は、評価の結果、既存の Wiki に比べ、検索性能の指標である F 値において、31.7%の性能向上を果たしていることが分かった。これについては評価の項で詳しく報告する。また、本利便性の高い Semantic Wiki のモデルの提案にあたり、このモデルで利用する Semantic Wiki におけるメタデータの自動生成手法とメタデータを活用したランキング手法も併せて提案する。この Semantic Wiki におけるメタデータの自動生成手法は、Wiki ページに対してメタデータを透過的に付与できる仕組みの提供を狙ったものである。また、メタデータを活用したランキング手法は、Wiki ページに付与されたメタデータを対象とした検索にて、検索結果の上位に有用な情報を出現させることを狙ったものである。評価では、これらの両手法についても、その評価結果を示す。

2. 関連研究と本研究との違い

Semantic Wiki とは、前述の通り、Semantic Web の技術を用いた Wiki の総称である。この Semantic Wiki では、Semantic Web の知識を持つ利用者が、Wiki ページに対し、RDF に基づき Dublin Core[5]などに準拠したメタデータを付与することが可能である。また、通常、メタデータ記述へのインセンティブとして、メタデータを用いた高度な検索機能が提供されている。

既存の Semantic Wiki として、PlatypusWiki[6][7]や SemperWiki[8][9]など様々な実装系が存在している。PlatypusWiki では、プルダウンメニューからタグを選び、値を入力するといった操作により、Wiki ページに対するメタデータの記述が行える。また、RDF トリプルを用いた検索機能も提供されている。SemperWiki は、個人利用を想定した Semantic Wiki であり、RDF トリプルに似た簡単な構文を Wiki ページの本文内に記述することで、メタデータの記述を行うことが可能である。また、記述されたメタデータを利用して、利用者が閲覧している Wiki ページに関連した Wiki ページの提示機能や、メタデータのタグと値を用いた検索機能を提供している。

しかし、これらの既存の Semantic Wiki では、共通して、以下の問題点を持っている。

- (a) メタデータの記述の際に、Semantic Web に関する専門知識が必要であり、また、メタデータの値となる情報を与えなければならないといった煩雑な手間も必要
- (b) メタデータを用いた検索を行う際に、検索クエリ記述のための Semantic Web に関する専門知識が必要
- (c) 検索を行った結果、多数の Wiki ページが候補としてあった場合、どの Wiki ページが適切であるか、Wiki ページの内容を逐次見て判断しなければならない

これらの問題点を解決するため、本研究の利便性の高い Semantic Wiki では、以下の3つの要件に対応する。

- (1) メタデータの生成に Semantic Web の専門知識や煩雑な手間を必要としない

† 岩手県立大学ソフトウェア情報学部

‡ 株式会社日立東日本ソリューションズ

(2) メタデータを用いた検索を行う際に、検索クエリを記述するための専門知識や煩雑な手間を必要としない

(3) 検索結果に候補となる Wiki ページが多数あった場合に、どの Wiki ページが適切であるかを順位付けて提示する

本研究では、上述の要件(1)を達成するために、Semantic Wiki におけるメタデータの自動生成手法を考案する。そして、Wiki ページの生成、編集を行った際に、Wiki ページに関する情報を解析し、この自動生成手法を用いて Dublin Core に準拠したメタデータの自動生成を試みる。

このようなメタデータの自動生成手法を考える上では、Web ページを対象とした既存のメタデータの自動生成手法の流用について考慮すべきである。既存の Web ページに対するメタデータの自動生成手法としては、メタデータの自動生成ツール DC-dot[10]で利用されている手法が知られている。DC-dot では、Web ブラウザを通して、Web ページの URL を入力すると、その Web ページに対し Dublin Core 準拠のメタデータを生成することが可能である。しかし、DC-dot は、指定された Web ページの文章構造を解析し、一部分の情報のみを用いてメタデータを生成するため、生成されるメタデータのタグ数は少なく(5項目程度)、また、精度も低い。このため、これをそのまま流用するのは不適切である。そこで本研究では、Wiki に特化した、タグ数が多く、精度も高いメタデータの自動生成手法を新たに考案することとした。

前述の要件(2)を達成するために、本研究では、通常の Wiki と同等の全文検索機能インタフェースを利用者に提供し、一方内部では、メタデータを対象として検索を行う。また、この際、前述の要件(3)を達成するために、Semantic Wiki におけるメタデータを活用したランキング手法を考案し、検索結果の上位に有用な情報が出現するよう試みる。

3. 利便性の高い Semantic Wiki

3.1 Semantic Wiki におけるメタデータの自動生成手法の提案

本研究で提案する Wiki に特化した、タグ数が多く、精度も高いメタデータの自動生成手法を表1に示す。

表1 Wiki ページからのメタデータ自動生成手法

タグ	取得する値
dc:title	Wiki ページのタイトル
dc:creator	Wiki ページの作成者、編集者 (取得できない場合は anonymous)
dc:subject	Wiki ページのタイトルから階層構造を抜いたもの、及び、リンク元 Wiki ページのエイリアス
dc:description	Wiki ページ本文の先頭一文
dc:publisher	サーバのドメイン
dc:contributor	Wiki サイト名
dc:date	Wiki ページの編集日
dc:type	dcmitype:Text
dc:format	text/html
dc:identifier	Wiki ページの URL
dc:source	内部リンク元の Wiki ページのリスト
dc:language	Wiki の設定言語
dc:relation	他の Web リソースへのリンクのリスト
dcterms:tableOfContents	Wiki ページの見出し

表1のうち、特に dc:subject と dc:description の値が特に Wiki に特化した部分となっている。

dc:subject では、Wiki ページのタイトルを取得しているが、タイトルが階層構造を表現したものであった場合は、より実体を表現している階層構造の最下部を取得するよう工夫している。さらに、その Wiki ページにリンクを張っている Wiki ページにおいて、エイリアスと呼ばれる別名リンクを使っていた場合、このエイリアスを Wiki ページに対する特徴的な単語として、dc:subject に取得する。

また、Wiki においては、先頭一文にその Wiki ページに対する簡単な説明を記述していることが多いため、dc:description では、本文の先頭一文を取得している。

3.2 Semantic Wiki におけるメタデータを活用したランキング手法の提案

本提案のランキング手法を検索の流れとともに以下に示す。

- (1) 利用者からの検索要求として検索語を受け取り、メタデータに対し全文検索を行う
- (2) 検索結果内の各メタデータに対して以下の処理を行う
 - (2-1) 検索語とメタデータ内の特定のタグの値と比較
 - (2-2) 一致した場合はそのタグの種類に応じて重みを加算
- (3) メタデータの得点の高い順に Wiki ページを並び替え提示する

本ランキング手法で利用するタグと、各タグに対する重みを表2に示す。

表2 利用するタグと各タグに対するする重み

タグ	完全一致時の重み	部分一致時の重み
dc:title	1	0.2
dc:subject	0.8	0.2
dc:description	0.4	0.1
dc:source	0.1	0.05
dc:relation	0.1	0.05
dcterms:tableOfContents	0.4	0.1

表2では、dc:title や dc:subject に特に高い重みを与えているが、これは、dc:title が最も Wiki ページの特徴を現しており、また、dc:subject には、Wiki ページを特徴付ける単語が出現しやすいためである。

本研究では、表1の14種のタグのうち表2に示す6種のタグのみ重みを与えている。これは、事前調査(25語のキーワードを用い、前述のメタデータの自動生成手法によって自動生成されたメタデータの各タグに対して検索後、検索結果の適合率、再現率、F値を検証)の結果、表2に示す6種のタグが Wiki ページの検索にて特に有用であったためである。

また、これら6種のタグに与えた重みの値については、評価にて、最も良い性能を得られたときの値とした。

3.3 利便性の高い Semantic Wiki のモデルの提案

我々の提案するメタデータの自動生成手法と、メタデータを活用したランキング手法を考慮した利便性の高い Semantic Wiki のモデルを図1に示し、図1の構成要素の詳細を以下に述べる。

Wiki: 既存の Wiki を利用し、Wiki ページが編集された際に、メタデータ生成部へ処理を渡す。また、メタデータ表示・編集機能と、メタデータ検索機能を提供するため、インタフェースを一部変更する。

メタデータ生成部: 我々の提案するメタデータの自動生成手法を用いてメタデータを自動生成し、メタデータデータベースに格納する。

メタデータ表示・編集部: メタデータ表示部では、利用者が現在閲覧している Wiki ページのメタデータの表示を行う。メタデータ編集部は、利用者が不完全、または間違っただメタデータを発見したときに補助的に利用されるものであり、メタデータの編集が行われた際に、編集されたメタデータをメタデータデータベースに格納する。本メタデータ編集部は、不十分な部分があれば誰かが補完するという Wiki の思想を引き継いだものとなっている。

メタデータ検索部: 我々の提案するメタデータを活用したランキング手法を用いて Wiki ページの検索を行い、検索結果の Wiki ページのリストを利用者に提示する。

メタデータデータベース: メタデータを格納する。

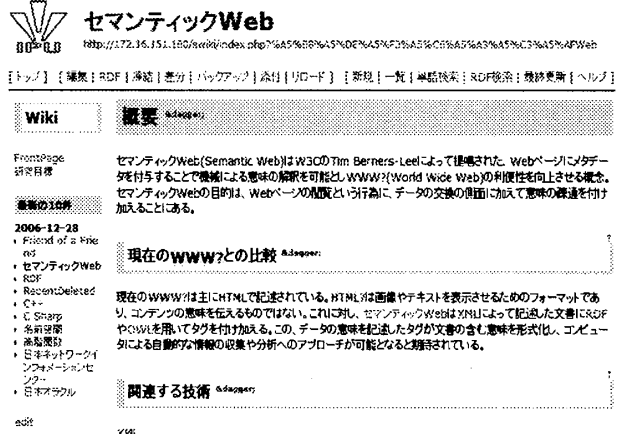


図3 本 Semantic Wiki のインタフェース

図3に示すように、PukiWikiのナビゲーションバーにメタデータ表示・編集部、メタデータ検索部へのリンクを追加した。また、Wiki ページの更新の際に、メタデータ生成部を呼び出すように処理を追加した。

本構築実験では、実験用に、コンピュータやインターネットに関する Wiki ページを1000件程度格納している。

5. 評価

本評価では、以下の項目について評価を行った。

メタデータの精度

本提案のメタデータの自動生成手法の有効性を確認するため、本提案手法によって生成されたメタデータの精度の評価を行った。

評価方法としては、構築実験環境で生成されたメタデータ10件に対し、コンピュータやインターネットについての知識のある被験者10人を用い、タグの値の内容に関する正誤判定を実施した。本評価結果(抜粋)を表3に示す。

表3 メタデータの精度の評価結果(抜粋)

		メタデータ							
		1	2	3	4	5	6	7	8
		メタデータ内に含まれていた項目数							
		12	18	19	20	22	26	23	23
被験者	A	12	18	17	19	22	26	23	22
	B	12	18	18	19	22	26	23	21
	C	12	18	19	20	22	26	23	23
	D	11	18	19	20	22	26	22	23
	E	12	18	19	19	21	25	21	22
	F	12	18	17	19	21	26	23	20
	G	12	18	19	20	22	26	23	23
	H	12	18	19	19	21	22	22	23
	I	12	18	16	19	22	26	21	21
	J	12	18	18	19	22	25	21	21
正解平均		11.9	18	18.1	19.3	21.7	25.4	22.2	21.9
精度 (%)		99.2	100	95.3	96.5	98.6	97.7	95.2	97.7

表3の結果から、メタデータに記述されているタグのうち、約97%のタグの値が適切と判断されており、本提案のメタデータの自動生成手法の有用性がうかがえる。

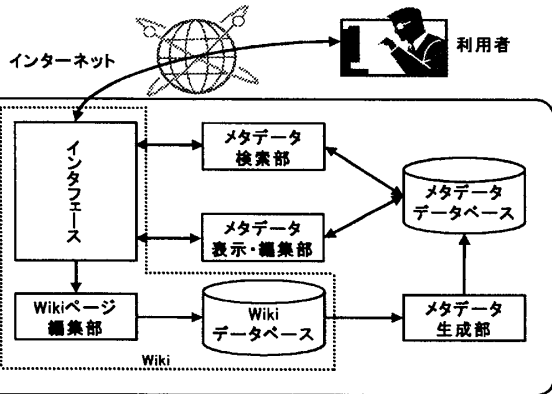


図1 利便性の高い Semantic Wiki のモデル

4. 構築実験

我々の提案する利便性の高い Semantic Wiki のモデルに従い、メタデータの自動生成手法とメタデータを活用したランキング手法を考慮した Semantic Wiki の構築実験を行った。図2に本構築実験環境を示す。

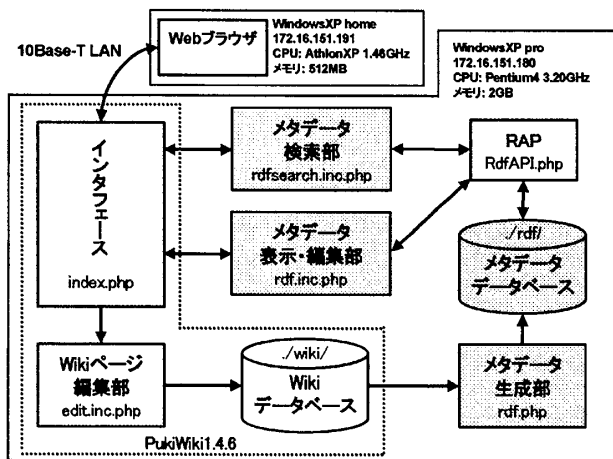


図2 本構築実験環境

本構築実験では、本 Semantic Wiki を1台のコンピュータ上に実装し、Web上に公開を行った。

プログラムの開発言語には PHP5 を使用した。ベースとなる Wiki としては、PukiWiki1.4.6[11]を利用している。RAP (RDF API for PHP)としては、RAPバージョン0.9.4[13]を SPARQL を用いた値の取得のために利用した。本 Semantic Wiki のインタフェースを図3に示す。

ランキング性能

本提案のメタデータを活用したランキング手法の評価として、JavaやWeblogといった、コンピュータに関連する10語を検索語として用い、検索語に対する正解を決定した上で、上位5件の適合率、上位5件の再現率(全正解のうち検索結果上位5件に含まれる正解の割合)、それらのF値、正解の出現位置の計測を行った。

また、メタデータを活用したランキング手法で用いている重みの正当性検証のため、全ての重みを0.1に設定した場合(パターン1)、パターン1でdc:titleとdc:subjectに高い重み(0.4)を設定した場合(パターン2)、パターン1でdc:descriptionとdcterms:tableOfContentsに高い重み(0.4)を設定した場合(パターン3)、パターン1でdc:relationとdc:sourceに高い重み(0.4)を設定した場合(パターン4)に分けて、重みの与え方による影響の確認を行った。

上位5件の適合率、上位5件の再現率、それらのF値、正解の出現位置の計測結果を表4に示す。

表4 ランキング性能

検索語	適合率 (上位5件)	再現率 (上位5件)	F値	正解の 位置
Java	20.0	100	33.3	1
Weblog	50.0	50.0	50.0	1
C#	33.3	100	50.0	1
HTML	20.0	100	33.3	1
Wiki	40.0	66.7	50.0	1,2
メタデータ	20.0	100	33.3	1
マークアップ言語	100	100	100	1,2,3,4
LAN	40.0	100	57.1	1,5
DNS	75.0	100	85.7	1,3,4
P2P	66.7	100	80.0	1,2
平均	46.5	91.7	57.3	

表4の結果より、本提案のメタデータを活用したランキング手法を用いた場合、検索結果の上位に正解となるWikiページが出現していることが確認できる。

また、上述の10語に対し、パターン1、パターン2、パターン3、パターン4の重みで、実際に検索を行った結果(平均値)を表5に示す。

表5 重みの与え方による影響(平均値)

検索語	適合率 (上位5件)	再現率 (上位5件)	F値
本提案の重み	46.5	91.7	57.3
パターン1	40.5	80.0	49.4
パターン2	42.5	83.3	51.9
パターン3	40.5	80.0	49.4
パターン4	38.5	67.5	40.6

表5より、本提案のメタデータを活用したランキング手法で用いている重みの有効性が確認できる。

検索性能

利便性の高いSemantic Wikiのモデルの有効性確認のため、本構築実験環境のSemantic Wikiと、既存のWiki(PukiWiki)との検索精度(適合率、再現率、F値)に関する比較評価を行った。検索に用いた検索語は、前述の10語である。本評価結果を表6に示す。

表6の結果から、再現率はほぼ同等であり、適合率、F値においては、改善されていることが分かる。

表6 検索性能の比較

検索語	既存の Wiki			本 Semantic Wiki		
	適合率	再現率	F値	適合率	再現率	F値
Java	0.9	100	1.7	7.1	100	13.3
Weblog	13.3	100	23.5	50.0	50	50.0
C#	5.3	100	10.0	33.3	100	50.0
HTML	0.3	100	0.6	8.3	100	15.4
Wiki	2.8	100	5.4	11.1	100	13.3
メタデータ	7.1	100	13.3	20.0	100	33.3
マークアップ言語	36.4	100	53.3	100	100	100
LAN	0.8	100	1.6	4.9	100	9.3
DNS	7.3	100	13.6	75.0	100	85.7
P2P	5.4	100	10.3	66.7	100	80.8
平均	8.0	100	13.3	37.6	95.0	45.0

6. おわりに

本研究では、本提案のSemantic Wikiにおけるメタデータの自動生成手法と、メタデータを活用したランキング手法の提案を行い、両提案手法を考慮した利便性の高いSemantic Wikiのモデルの提案を行った。

また、本提案の利便性の高いSemantic Wikiのモデルに従った構築実験について報告した。そして、本構築実験環境を用いて行った評価結果について説明し、本提案の有用性を示した。

今後の課題として、Webオントロジを利用した検索機能の強化が考えられる。

参考文献

- [1]Ward Cunningham: Wiki Wiki Web
<http://c2.com/cgi/wiki>
- [2]Tim Berners-Lee: The Semantic Web
<http://www.w3.org/2002/Talks/04-sweb/>
- [3]SemanticWeb.org
<http://www.semanticweb.org/>
- [4]神崎 正英: セマンティック・ウェブのための RDF/OWL 入門, 森北出版(2004).
- [5]Dublin Core Metadata Initiative (DCMI)
<http://dublincore.org/>
- [6]Roberto Tazzoli, Paolo Castagna, Stefano Emilio Campanini: Towards a Semantic WikiWiki-Web, Proc. of 3rd International Semantic Web Conference, pp.7--11 (2004).
- [7]PlatypusWiki
<http://platypuswiki.sourceforge.net/>
- [8]Eyal Oren, John G. Breslin, Stefan Decker: How semantics make better wikis, Proc. of 15th International World Wide Web Conference, pp.1071--1072 (2006).
- [9]SemperWiki
<http://www.semperwiki.org/>
- [10]DC-dot
<http://www.ukoln.ac.uk/metadata/dcdot/>
- [11]PukiWiki 公式サイト
<http://pukiwiki.sourceforge.jp/>
- [12]SPARQL Query Language for RDF
<http://www.w3.org/TR/rdf-sparql-query/>
- [13]RAP - RDF API for PHP V0.9.4
<http://sites.wiwiss.fu-berlin.de/suhl/bizer/rdfapi/>