

## Web 上の Web カメラ検索環境のモデルの 提案とその評価

劉 海燕<sup>†</sup> 児玉英一郎<sup>†</sup> 王 家宏<sup>†</sup> 高田豊雄<sup>†</sup>

近年、Web 上での Web カメラの利用が進んでいる。このため、Web 上に発行された Web カメラ数は年々増加しており、必要な Web カメラの検索が重要なものとなってきている。このような状況の中、いくつかの Web カメラ検索の専用サイトが登場している。しかし、これら既存のサイトでは、Web カメラ数を増やす点、メタデータを利用した高度な検索機能の提供の点で課題を抱えている。本研究では、この課題を解決し、利用者が目的とする Web カメラを Web 上から時間的及び精神的負荷なく検索できる環境のモデルを提案する。また、本環境のプロトタイプを利用した評価についても報告する。

### Proposal and Performance Evaluation of Searching Model of Web Cameras on Web

Haiyan Liu<sup>†</sup> Eiichiro Kodama<sup>†</sup> Jiahong Wang<sup>†</sup>  
and Toyoo Takata<sup>†</sup>

In recent years, web cameras have become more and more widely used, the number of Web cameras increases from year to year, and the searching method for finding Web cameras becomes important. In response to this requirement, many searching sites have been provided for use. It is still difficult, however, to adopt a new web camera into an existing site or provide advanced searching functions based on the metadata. For solving these problems, we propose a searching model that can help users find web cameras easily with little mental strain and without costing too much time. We have studied performance of the proposed model by implementing a prototype system and conducting

extensive experiments. This paper reports results of the performance study.

#### 1. はじめに

##### 1.1 背景

近年のインターネットの普及に伴い、インターネットは多種多様な使われ方をするようになった。例えば、その1つに Web カメラを使用したインターネットの利用方法が挙げられる。これは、観光地などにあるホテルの屋上に Web カメラを設置し、利用者に現地の最新の情報を提供するものであり、観光地である安比高原や富士山などの現地情報が実際に提供されている。これにより、安比高原にスキーに行きたい人などが、当日の雪の状況や詳しい天候などを目視確認でき、利用者にとっては便利なものとなっている。

しかし、このような恩恵を利用者が得るためには、サーチエンジンなどを利用して、利用者の目的に適った Web カメラを発見することが必要となる。検索対象がはっきりしているような状況では、例えば、安比高原の Web カメラを見つけないといった場合、安比高原をキーワードとして、サーチエンジンを利用することにより、安比高原の Web カメラを見つけることが可能である。しかし、検索対象が抽象的な場合には、例えば、ゲレンデがきれいなスキー場の Web カメラを見つけないといった場合、サーチエンジンでは検索が困難となっている。

このような問題への対応として、最近では、カメ探、ライブカムジャパン、世界の窓といった Web カメラ検索の専用サイトが利用されている。

##### 1.2 既存の Web カメラ検索の専用サイトの問題点

本研究の考察対象である、カメ探[1]、ライブカムジャパン[2]、世界の窓[3]といった既存の Web カメラ検索の専用サイトの抱える問題点を以下に示す。

- サイト管理者が Web カメラを人手により登録している  
Web カメラ検索の専用サイトが検索の対象とする Web カメラ総数は、網羅性確保の観点から重要なものとなっている。しかし、世界の窓などの既存の Web カメラ検索の専用サイトが検索の対象とする Web カメラは、サイト管理者が人手で登録しており、この方式が Web カメラ数を増やす際の障害となっている。
- メタデータを利用した高度な検索機能が不足している  
既存の Web カメラ検索の専用サイトでは、全文検索が主流となっており、メタ

<sup>†</sup> 岩手県立大学大学院ソフトウェア情報学研究科  
Graduate School of Software and Information Science, Iwate Prefectural University

データに対応した高度な検索機能を提供していない。一部の Web カメラ検索の専用サイトでは、メタデータを利用した検索機能を提供しているが、検索時に指定可能なメタデータの項目数が少なく、利便性の点からは不十分なものとなっている。

### 1.3 目的

本研究では、Web 上から多くの Web カメラを自動的に収集し、これにメタデータを半自動的に付与してデータベースに蓄積するとともに、これを利用して、利用者にとって検索しやすい Web カメラの検索環境を構築することを目的としている。

## 2. 関連研究

### 2.1 Web カメラの収集に関する関連研究

本研究の目的達成のためには、Web 上から Web カメラを検出し自動的に収集する必要があるが、Web カメラを直接収集してくるようなクローラに関する研究は知られていない。このため、関連するものとしては、一般の Web ページを対象としたクローラに関する研究となってしまう。一般の Web ページを対象としたクローラに関する研究は、1994 年頃から数多く行われており、例えば、B. Pinkerton[4]による研究が挙げられる。

### 2.2 メタデータに関する関連研究

本研究では、メタデータを利用した高度な検索機能の提供を試みるが、メタデータの利用についての関連研究として、動画を対象にメタデータを付与し、これを利用して検索を行う研究[5]が知られている。この研究では、動画をに対し内容解析を行い、色合いなどに関する独自のメタデータを作成している。本研究は、Web カメラを対象にしており、かつ、Web カメラの埋め込まれた Web ページのテキスト情報に対する解析結果を中心的に利用し、Dublin Core に準拠したメタデータを作成するものであり、メタデータを生成し、利用するといった点以外は異なるものとなっている。

また、本研究では、メタデータの半自動生成を試みているが、メタデータの半自動生成に関しては、メタデータの生成支援ツールが関連する。このメタデータの生成ツールとしては、The Web 神崎のメタデータ生成ツール[6]や、古典的な dcdot[7]などが知られている。

本研究では、Dublin Core に準拠したメタデータの利用を想定しているが、このメタデータの標準に関しては、セマンティック Web などにおけるメタデータの標準である Dublin Core、学習用メタデータの標準である LOM (Learning Object Metadata)、EU 電子政府用メタデータの標準である MIREG (Management Information Resources for e-Government) など様々な標準が存在しており、分野に応じて使い分けられている状況となっている。本研究では、Web 上の Web リソースである Web カメラを対象としていることから、Dublin Core を利用している。

### 2.3 Web カメラの利用に関する研究

Web カメラの利用に関する関連研究として、酒匂らのメタデータを利用した Web カメラ検索環境の構築に関する研究[8]が挙げられる。この研究では、Web カメラ公開閲覧専用サイト内に発行された Web カメラを、メタデータを利用して効率的に検索することが可能である。Web カメラを対象にメタデータを生成し、検索に利用するといったアプローチの点で、本研究に深く関連している。しかし、Web 上に点在する Web カメラを収集対象としている点、メタデータの生成方法の点で本研究とは異なっている。

## 3. Web カメラ検索環境のモデルの提案

### 3.1 Web カメラ検索環境のモデルの全体像

本研究で提案する Web カメラ検索環境のモデルを図 1 に示す。

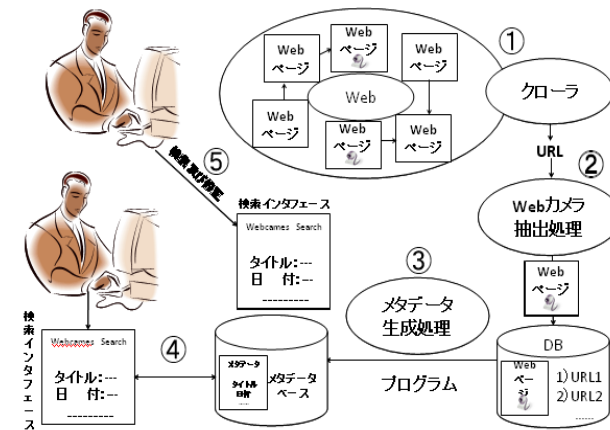


図1 Web カメラ検索環境のモデル

本提案モデルの動きは以下のようになっている。

- ①：クローラにより、Web 上から Web ページ群の収集を行う。
- ②：Web ページ群から、後述の Web カメラ抽出手法により、Web カメラを公開している Web ページのみを抽出し、データベースに蓄積する。
- ③：抽出された各 URL に対し、本研究独自のメタデータ生成手法により、メタデータを半自動的に生成後、メタデータベースに蓄積する。
- ④：利用者は、メタデータを対象とした検索インタフェースを利用して検索を行う。
- ⑤：また、メタデータに不備がある場合、利用者がその修正を行う。

### 3.2 Web カメラ抽出手法

本提案モデルにおいては、Web カメラを公開している Web ページの収集が重要である。以下、本研究で新たに考案した Web カメラ抽出手法を示す。

A を Web ページのソースの title タグに Web カメラに関するキーワード群 (Web カメラ, ライブカメラ, ウェブカメラ, ライブカム, Webcams, Webcam, Live camera) 内の単語を含んでいる Web ページの集合とする。

そして、B は Web ページのソースのメタタグ内の keyword 属性に前述の Web カメラに関するキーワード群内の単語を含んでいる Web ページの集合とする。

また、C は A ∪ B 内の Web ページで、Web ページ内の a タグのアンカーテキストに製品に関するキーワード群 (製品, 商品, 価格) 内の単語を含んでいるものの集合とする。(但し、記号-は差集合を表す。)

このとき、集合 A ∪ B - C を Web カメラを含んでいる Web ページの集合とする。

### 3.3 メタデータ生成手法

本研究では、セマンティック Web でよく利用されている Dublin Core の標準の 15 項目を利用しメタデータを生成する。以下、本研究で新たに考案したメタデータ生成手法を示す。

- dc: title の値の取得：  
本値は Web ページソース内の title タグの値から取得する。
- dc: description の値の取得：  
本値は Web ページソース内に description 属性を持つ meta タグがあれば、その値から取得する。
- dc: creator の値の取得：  
本値は Web ページソース内に author 属性を持つ meta タグがあれば、その値から取得する。
- dc: language の値の取得：  
本値は Web ページソース内の html タグに lang 属性がある場合には lang 属性の値から取得する。ない場合には、title タグや meta タグの keyword 属性に Web カメラに関するキーワード群内の特定の単語 (Web カメラ, ライブカメラ, ウェブカメラ, ライブカム) を一つでも含んでいれば、ja に設定する。
- dc: subject の値の取得：  
本値は Web ページソース内の meta タグの keyword 属性の値から取得する。
- dc: right の値の取得：  
本値は Web ページソース内のテキスト部分に copyright の単語が出現した場合には、copyright の後ろの単語を値として取得し、ない場合には、Web ページの URL 内のホスト名にする。
- dc: identifier の値の取得：

本値は Web カメラを公開している Web ページの URL とする。

- dc: date の値の取得：  
本値は http ヘッダ内の MIME タイプ date の last modified の値から取得する。

## 4. Web カメラ検索環境の構築実験

### 4.1 実験環境

サーバコンピュータ 1 台, クライアントコンピュータ 1 台の計 2 台を用い, Web カメラ検索環境の構築実験を行った。

サーバ側には, Web カメラを含む Web ページの URL を格納するためのデータベースとして SQL Server2005 をインストールし, また, メタデータ格納用のデータベースとして RDF データベースの Sesame2.0 をインストールした。

また, Web ページを収集し本研究独自の Web カメラ抽出処理を行うクローラ (Spider.sln), 本研究独自のメタデータ生成処理をするプログラム (AddRDFFiles.java) の実装も行った。そして, ユーザインタフェース構築のためサーブレットコンテナとして Jetty6.1 をインストールし, サーブレットとしてユーザインタフェース部分 (Search WebCam.jsp) の実装を行った。

本構築実験により構築された環境を図 2 に示す。

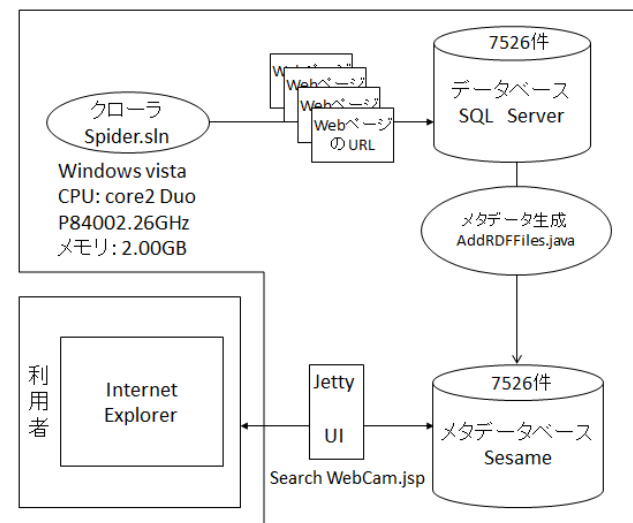


図 2 構築実験により構築された環境

## 4.2 構成要素

以下、図2の各構成要素の詳細を順に示す。

- クローラ

本クローラは Web 上から Web ページ群を収集し、本研究の独自の Web カメラの抽出手法により、Web カメラを公開している Web ページを抽出するプログラムである。

本クローラ及び Web カメラの抽出処理部分は C# で実装を行った。図3に本クローラのインタフェースを示す。

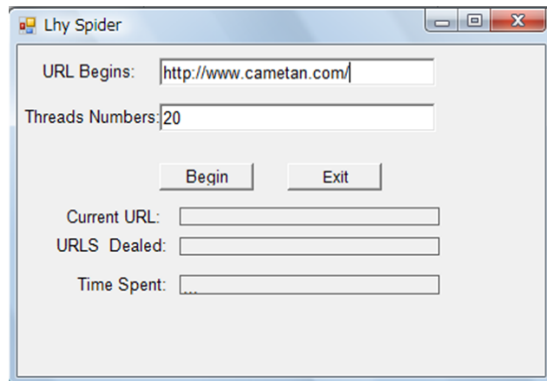


図3 クローラのインタフェース

- データベース

前述のクローラにより抽出された Web カメラを公開している Web ページの URL を蓄積している。現在のところ 7526 件の URL を格納している。

- メタデータベース

本研究独自のメタデータ生成手法により、抽出された各 URL に対し、メタデータの自動生成を行い本メタデータベースに格納している。本データベースも、現在のところ 7526 件のデータを格納している。

- 検索インタフェース

利用者にメタデータを対象とした検索インタフェースを提供する。本検索インタフェースを図4に示す。



図4 検索インタフェース

## 5. 評価

### 5.1 評価内容

本評価においては、基本性能評価と本提案内容の評価を行う。以下に、評価項目を示す。

#### 5.1.1 基本性能評価

基本性能評価においては、適合率を求め既存 Web カメラ検索サイトと比較する。

#### 5.1.2 提案内容の評価

本提案内容についての評価として、以下の(1)～(3)について評価を実施する。

- (1) Web カメラ数の比較
- (2) Web カメラ抽出手法の精度検証
- (3) メタデータの生成手法の精度検証

### 5.2 評価方法

前述の評価項目に対して、次の評価方法に従い評価を行った。

#### 5.2.1 基本性能評価

被験者 1 人に対し、岩手山、盛岡、伊豆、東京タワー、銀座の 5 つのキーワードを用い、既存の Web カメラ検索の専用サイトである世界の窓と本システムとの適合率の比較を行った。

#### 5.2.2 提案内容の評価

- (1) Web カメラ数の比較：本研究の保有する Web カメラ数を既存の Web カメラ検索サイトと比較する。

(2) Web カメラ抽出手法の精度検証：

Web カメラを含んでいると判定した Web ページ群からサンプルを抽出し、そのサンプルに対し、Web カメラが含まれていたか、人手により判定する。サンプル数  $n$  の決定は、必要サンプル数を求めるための公式  $n = Z^2 \sigma^2 / E^2$  [9] により求めた。但し、 $Z$  は誤差の信頼率（信頼度 95% の場合、 $Z$  の値は 1.96）、 $\sigma$  は標本の分散（ここでは、 $\sigma = 0.5$  とした）、 $E$  は要求精度（今回は 5% の誤差を許容、 $E = 0.05$  とした）である。

以上の数値により計算した結果、 $n = 385$  となった。したがって、今回はサンプル数を 400 とした。

(3) メタデータの生成手法の精度検証：

3 名の被験者を用い、10 個の Web カメラを含んでいる Web ページに対し、メタデータの正解集合を作成後、自動生成されたメタデータの各項目が正解かどうかを手により判定する。

5.3 評価結果

5.3.1 基本性能の評価結果

被験者を用い実施した本評価結果を表 1 に示す。

表 1 世界の窓と本システムの適合率の比較

検索キーワード	世界の窓	本システム
岩手山	1/3 (33.3%)	1/1 (100%)
盛岡	2/3 (66.7%)	2/2 (100%)
伊豆	3/7 (42.9%)	2/2 (100%)
東京タワー	1/3 (33.3%)	1/1 (100%)
銀座	1/3 (33.3%)	1/1 (100%)

5.3.2 提案内容の評価

(1) Web カメラ数について比較した結果を表 2 に示す。

表 2 Web カメラ数の比較結果

Web カメラ検索の専門サイト	保存数
カメ探	1207
ライブカムジャパン	約 2000 (概算)
世界の窓	2394
本研究	約 7037 (7526×0.935)

(2) Web カメラ抽出手法の精度検証

Web カメラを含んでいると判定した Web ページ群 7526 件からサンプル 400 件を抽出し、そのサンプルに対し、Web カメラが含まれていたか人手で判定した結果、93.5% (約 7037=7526×0.935) の Web ページにおいて Web カメラが含まれていた。

(3) メタデータの生成手法の精度検証

被験者 3 名により精度検証を行った結果を表 3、表 4、表 5 に示す。

表 3 被験者 1 により精度検証した結果

10 件	生成率	正解率
dc:title	100%	80%
dc:description	60%	50%
dc:creator	0%	0%
dc:subject	60%	83.3%
dc:identifier	100%	100%
dc:language	100%	100%
dc:right	40%	75%

表 3 の正解率のうち、生成率が 0% であった dc: creator の部分を除いたものの平均は 81.4% となった。

表 4 被験者 2 により精度検証した結果

10 件	生成率	正解率
dc:title	100%	80%
dc:description	60%	50%
dc:creator	0%	0%
dc:subject	60%	83.3%
dc:identifier	100%	100%
dc:language	100%	100%
dc:right	40%	100%

表 4 の正解率のうち、dc: creator の部分を除いたものの平均は 85.6% となった。

表 5 被験者 3 により精度検証した結果

10 件	生成率	正解率
dc:title	100%	70%
dc:description	60%	66.7%
dc:creator	0%	0%
dc:subject	60%	50%
dc:identifier	100%	100%
dc:language	100%	100%
dc:right	40%	100%

表 5 の正解率のうち、dc: creator の部分を除いたものの平均は 81.1% となった。

## 6. まとめ

本研究では、既存の Web カメラ検索の専用サイトにおける問題点を解決する、利便性の高い Web カメラ検索環境のモデルを提案した。また、本提案モデル内で利用する本研究の独自の Web カメラの抽出手法、及び、メタデータの生成手法の提案も行った。

また、本モデルに従い、構築実験としてプロトタイプを作成を行った。そして、本プロトタイプを利用した評価を行い、その結果を示した。

## 参考文献

- [1] <http://www.cametan.com/>
- [2] <http://orange.zero.jp/zad23743.oak/livecam/>
- [3] <http://www.sekainomado.com/>
- [4] B. Pinkerton: Finding what people want: Experiences with the WebCrawler, Proceedings of the 2nd International World Wide Web Conference, pp.7--18 (1994).
- [5] 上野太一, 倉林修一, 清木康: 動画像における色彩特徴量の分析による感性メタデータ自動生成及び時系列メディアデータ検索機構の実現, 情報処理学会研究報告, 2008-DBS-146, pp. 349--354 (2008).
- [6] <http://www.kanzaki.com/docs/sw/dc-a-matic>
- [7] <http://www.ukoln.ac.uk/metadata/dcdot/>
- [8] 酒匂 大輔, 児玉 英一郎, 王家宏, 高田 豊雄: 次世代ネットワークにおける

Web カメラの検索に関する考察, 平成 20 年度 電気関係学会東北支部連合大会 講演論文集, p.219 (2008).

[9] <http://wenku.baidu.com/view/79f14ac52cc58bd63186bd9c.html>