

多様な Web 機器に対応可能なデータ収集フレームワークの提案

北島 啓太郎[†] 橋 拓馬^{**} 峰野 博史[†]

[†] 静岡大学情報学部 ^{**} 静岡大学大学院総合科学技術研究科

1. はじめに

近年, IoTの普及に伴い無線LAN・HTTPブラウザ経由で操作可能な機器 (Web機器) が増加している[1,2]. 一方, Web機器が増えると同時に, Web機器自身が無線LAN親機となるネットワーク構成も一般的になりつつある. Web機器自身が無線LANの親機となる構成の場合, スマートフォンなどのクライアントが無線LAN子機として機能する. そのため複数のWeb機器に対して通信処理を行なうことについて想定されてないといった課題が存在する. 例えば無線LAN機能付カメラの場合, 撮影した動画をHTTPを用いてスマートフォンに送信する機能が実装されているものの, 複数の無線LAN親機にスマートフォンは同時接続を行なうことはできない. そのため, スマートフォンが複数台の無線LAN機能付カメラを操作したい場合は手動にて無線LAN接続を切り替える必要があり, 切り替える度に操作コストが掛かってしまっている.

本研究では, Web機器が無線LANの親機となる構成において複数のWeb機器に対する無線LAN接続の切替コストが存在する課題を解決するため, 複数のWeb機器への処理を考慮し, 多様なWeb機器を一元的に扱うことの可能なデータ収集フレームワークを提案する. 提案するデータ収集フレームワークは, 多様なWeb機器に対する通信路切替, 複数Web機器の時刻同期, データ収集の重み付けを行なうための共通インターフェースを提供する.

2. 関連研究

多様な Web 機器のデータを収集する手法として, Web スクレイピングツールが挙げられる. Web スクレイピングツールの一例として Import.io[3], Nokogiri[4], Selenium[5]などがある.

Import.io は Import.io 社によって開発された Web スクレイピングサービスである. GUI を用いており, 誰でも容易にデータ収集が可能なツールであるが, Web サービスであるため, ローカルネットワークで機能する Web 機器には適用させることが不可能である. つまり Web 機器のデータ収集には不向きである.

Nokogiri は Ruby で動作するデータ収集フレームワークである. プログラムの追加実装が必要となるが, ローカルネットワーク内の Web 機器からデータ収集を行なうことができる. しかし, Nokogiri は Ruby 言語のみでしか動作せず, 更に無線 LAN の通信路切替の機能を実装する必要がある. つまり多様な Web 機器に対応可能なデータ収集フレームワークするには大幅な改良を要する.

Selenium は Web ブラウザの動作をシミュレートするラ

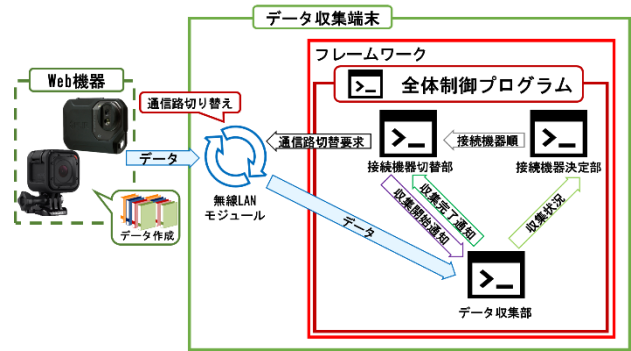


図 1: 提案フレームワーク全体像

イブラリである. 実際の Web ブラウザで行なう動作を自動化することが可能で, Java や Ruby, Python などの複数のプログラミング言語で動作させられる. しかし, Selenium の実行には実際のブラウザを動作させることが必要である上, Selenium のバージョンによって挙動が異なる, Selenium とブラウザの相性が存在するといった課題が存在する. Nokogiri と同様に無線 LAN の通信路切替機能を実装する必要がある. つまり多様な Web 機器に対応可能なデータ収集フレームワークとするには大幅な改良を要する.

以上のことから, 通信路切替の実装を必要としない既存の Web スクレイピングツールは現在存在しない. そのため, 通信路切替機能を別途実装する必要のないデータ収集フレームワークを提案する.

3. 提案手法

3.1 概要

本研究では, 通信路切替機能を別途実装する必要のないデータ収集フレームワークを提案する. 多様な Web 機器に対応可能なデータ収集フレームワークの全体像を図 1 に示す. フレームワークは“接続機器決定部”, “接続機器切替部”, “データ収集部”の 3 つのモジュールにて構成される. 接続機器決定部は Web 機器のデータ収集状況を基に, 接続する Web 機器の順番を決定する. 接続機器切替部では接続機器決定部にて定まった Web 機器の接続順番を基に Web 機器の接続切替処理を行なう. 接続切替処理はデータ収集部の動作が終了したことをトリガにして切替処理を開始する. 収集状況は接続機器決定部に順次送信する.

提案フレームワークの動作フローを図 2 に示す. データ収集プログラム開始時, 接続機器決定部は Web 機器の一覧情報を元に接続機器順リストを作成する. 接続機器順リストを元に, 接続機器切替部は Web 機器に対する接続機器切替をする. 接続機器切替後, データ収集部は Web 機器からデータ収集を行なう. データ収集の状況を接続機器決定部に通知することにより, 接続機器切替部にて接続機器順リストの接続順位を更新する. 更新された接続機器順リストを元に, 接続機器切替部は Web 機器

A proposal of a data collection framework for various Web Devices

Keitaro Kitajima[†], Takuma Tachibana^{**}, Hiroshi Mineno[†]

[†] Faculty of Informatics, Shizuoka University

^{**} Graduate School of Integrated Science and Technology, Shizuoka University

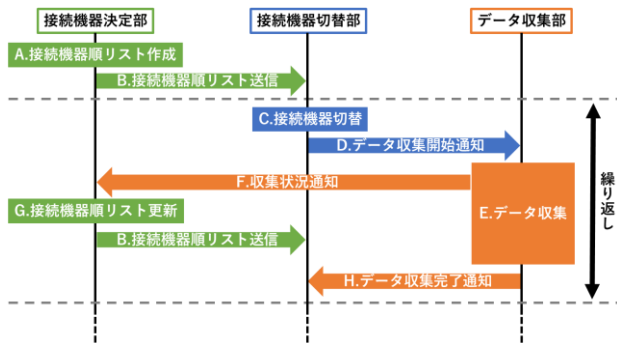


図2: 提案フレームワーク動作フロー

に対する接続機器切替をする。

3.2 接続機器決定部

データ収集プログラム開始時、接続機器決定部はデータ収集を行なう Web 機器の接続機器順位リストを作成する (図 2-A)。接続リストの一覧は予め用意した接続機器リストを利用する。取得した Web 機器の一覧情報を元に Web 機器に接続する順番を記した接続機器順位リストを作成し (図 2-A)、接続機器切替部に対して送信する (図 2-B)。接続リストの接続順位が更新される時は、データ収集部から収集状況の通知を受け取った時である (図 2-F)。データ収集部から送信された収集状況のメタ情報には“接続機器”、“データ生成日時”、“データ収集日時”が記録されている。メタ情報を元に、接続機器決定部では Web 機器の接続機器順位リストを更新する (図 2-G)。リスト更新後、接続機器決定部は接続機器切替部に対して接続順位を更新した接続機器順位リストを送信する (図 2-B)。

3.3 接続機器切替部

接続機器切替部は接続機器決定部から送信された接続機器順位リストを元にデータ収集端末の無線 LAN モジュールに対して通信路切替要求をする。通信路切替要求は、データ収集部にて Web 機器からデータ収集 (図 2-E) をしていないときに行なう。データ収集部が現在収集を行っているか確認するために、接続機器切替部内に収集中フラグを保持する。通信路切替完了後、接続機器切替部はデータ収集部に対してデータ収集開始通知を送信する。データ収集開始通知を出すと同時に、接続機器切替部は収集中フラグを上げる。収集中フラグを用いることにより、データ収集部にてデータ収集 (図 2-E) を行なっている間に別の Web 機器に接続を切り替えることを防ぐ。データ収集部よりデータ収集完了通知 (図 2-H) を受信したら収集中フラグを下げ、接続機器切替をする (図 2-C)。接続機器切替には、接続機器決定部より受信した接続機器順位リスト (図 2-B) を用いる。通信路切替順は接続機器順位リストの先頭から順次実行する。接続機器順位リストの最後まで接続機器切替を終えた際には、リストの初めから再度通信路切替要求をする。接続機器決定部から接続機器順位リストの更新があった際には、接続機器切替部内に保持したリストは破棄し、新しい接続機器切替順位リストの頭から再度通信路切替を行なう。

3.4 データ収集部

データ収集部は接続機器切替部から受信した収集開始通知を元に Web 機器からデータの収集を行なう。データ収集部は、接続機器切替部から送信されたデータ収集開

始通知 (図 2-D) をトリガにして処理を開始する。データ収集部はデータ収集開始通知に含まれた収集対象端末情報を元に無線 LAN を経由してデータ収集を行なう (図 2-E)。収集状況 (“接続機器”、“データ生成日時”、“データ収集日時”) は逐次接続機器決定部に送信する (図 2-F)。Web 機器の挙動は機器ごとに異なるため、データ収集 (図 2-E) と接続機器決定部への収集状況通知 (図 2-F) のみユーザの実装を必要とする。データ収集はシェルコマンドを経由して実行される。シェルコマンドを経由して実行されるため、データ収集フレームワークの実装とは異なるプログラミング言語でもデータ収集の実行が可能である。データ収集において要求される仕様は、Web 機器内のデータを収集可能なこと、接続機器決定部に収集状況の通知をすることの2つである。2つの条件を満たすプログラムをユーザは実装すればよいため、ユーザ側は既存手法にて述べた Nokogiri や Selenium といったローカルにて動作する Web スクレイピングツールを利用した開発が可能になる。データ収集が完了した際には接続機器切替部に対し、収集完了通知を送る (図 2-H)。収集完了通知をすることで、次の Web 機器に対して通信路の接続切替が行なわれる。

4. おわりに

本研究では、複数の Web 機器への処理を考慮し、多様な Web 機器を一元的に扱うことが可能なデータ収集フレームワークを提案した。今後はプロトタイプ実装と、実際の Web 機器を用いたデータ収集に提案するフレームワークを用いたプログラムを適用し、評価を予定している。プロトタイプ実装には無線 LAN に対応したカメラ “GoPro HERO5 Session” [6] を対象にすることを予定している。“GoPro HERO5 Session” を対象にしたプロトタイプ実装後は、他の Web 機器を対象にした実装、複数種類の Web 機器の併用が可能なフレームワークの実装を予定している。評価方針として、本手法を用いない場合とデータ生成からデータ収集までの遅延時間の比較を行なう。その後、実際の環境に適用して有用性をソフトウェアの可用性から評価する。

参考文献

- [1] “Cisco Visual Networking Index: Forecast and Methodology, 2016–2021”, Cisco, 2017.
- [2] Al-Fuqaha, A., Guizani, M., Mohammadi, M., et al.: Internet of things: A survey on enabling technologies, protocols, and applications., *Proc. IEEE Communications Surveys & Tutorials*, pp.2347-2376, IEEE (2015).
- [3] Import.io: Import.io (online), available from < <https://www.import.io> > (accessed 2018-01-11).
- [4] Team Sparklemotion: Nokogiri 鋸 (online), available from < <https://www.import.io> > (accessed 2018-01-09).
- [5] Selenium Project: Selenium WebDriver, Selenium (online), available from < [http:// www.seleniumhq.org/projects/webdriver/](http://www.seleniumhq.org/projects/webdriver/) > (accessed 2018-01-09)
- [6] Gopro, Inc. : GoPro - HERO5 Session 防水カメラ, GoPro 公式ウェブサイト (オンライン), 入手先 (< <https://jp.shop.gopro.com/APAC/cameras/hero5-session/CHDHS-502-master.html> >), (参照 2018-01-11)