

気象庁XML取得API 7年間の運用のまとめ その1

2020年10月1日

先端IT活用推進コンソーシアム
クラウド・テクノロジー活用部会 荒本道隆

■背景/活動目的

10年目は、量子アニーリングやAIについて学習を行い、その中でクラウドサービスを積極的に使用した。

11年目の最後の年は、**気象庁APIの運用で蓄積された7年半分のアクセスログとデータの、可視化やデータ活用について検討する。**

また、これまで通り、**クラウド全般**、IoT、量子コンピューティング、AIについても取り組んでいく。

広範囲の各要素技術に対し、部会参加者が「知っている」ではなく「使ったことがある」「人に教えることができる」と言えるレベルを目指す。

■活動方法

- ・月例オンラインミーティングの開催
- ・SNS 上での情報交換

■体制 (敬称略)

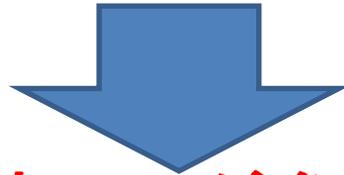
リーダー：荒本 道隆 (アドソル日進株式会社)

サブリーダー：上村 準也 (キヤノンITソリューションズ株式会社)

気象庁XML取得API: 作成した動機

2012年末に気象庁XMLのPUSH配信が試行されたけど....

- PUSH配信は、受信用サーバを立てないと受け取れないので、特に企業では「ちょっと受信してみよう」が難しい
- 気象庁のWebサイト上には過去1週間分しか履歴がない
(Feedが配信されて、気象庁Webサイト上のXMLをゲットする)
- 新規参入者が「去年の台風時のXMLを見たい」ができない



いつでも誰でも気象庁XMLが参照できるサーバを立てよう

- 気象庁Webサイトからコピーして、全履歴を置いておく
- Webブラウザを使って、誰でも簡単に参照できる
- AITCは非営利団体なので、ベストエフォート型
 - 2人でボランティアベースで7年半に渡って運用した

年表

- 2009年05月 「気象庁防災情報XML」仕様公開
- 2011年05月 「気象庁防災情報XML」運用開始
- 2012年12月 気象庁が、誰でも無料で受信できるPUSH配信試行を開始
- 2013年01月31日 クラウド部会有志が大崎のカフェで「いっちょAPIを作ってみるか」
- 02月09日 『気象庁XML取得API』をAWS上で運用開始 → 一般公開
- 08月 検索性能改善のため、事前集計処理を導入
- 09月 サーバ費が10倍(\$25→\$214)に。事前集計処理でDisk I/Oが激増
- 10月 性能改善し、事前集計処理を削除
- 11月 サーバ費が\$68/月に。今後もネットワークI/Oが増え続けるはず
- 12月 固定料金のさくらインターネットにお引越
- 2019年07月 気象庁XMLのサイズが巨大化。HDDが足りなくなる → 履歴を圧縮
- 2020年09月01日 気象庁XMLのPUSH配信が終了
- イマココ
- 2021年02月末 気象庁XML取得API停止予定。卒業研究に使う人が多いはず

- さくらVPSの仮想サーバ
 - 「メモリ2GByte、HDD-200GByte、CPU-3コア」プラン。「東京」
 - 月額1,580円 x 3
 - cloud1.aitc.jp [59.106.191.54] REST & Websocket
 - cloud2.aitc.jp [59.106.191.211] SPARQL
 - cloud3.aitc.jp [59.106.191.139] 動作検証 & 実験用
 - 備考
 - 6年半で3回ほど、メンテナンスのための再起動があった

<http://api.aitc.jp>



The screenshot shows a web browser window displaying the AITC API website. The browser's address bar shows 'api.aitc.jp'. The website header features the AITC logo and a '公開API' (Public API) button. A sidebar on the left contains a '公開API' menu with sub-items: 'はじめに', '気象庁XML用API', and '利用条件・留意事項'. The main content area has a '公開API' heading followed by three red text announcements regarding API service changes and extensions. Below these are sections for 'はじめに' and '気象庁XML用API', each with a brief introduction and a list of available API types.

公開API

はじめに
気象庁XML用API
利用条件・留意事項

公開API

2019/10/28 運用期間を2021年08月→2020年08月に変更しました。
気象庁からのPUSH配信が2020年8月末をもって終了するためです。ご了承ください。

2018/08/17 運用期間を2018年08月→2021年08月に延長しました。

2016/07/02 運用期間を2016年08月→2018年08月に延長しました。

■はじめに

AITC（先端IT活用推進コンソーシアム）の活動の中で作成したAPIを公開します。
ご利用いただくにあたっては、事前に利用条件・留意事項をご確認ください。

■気象庁XML用API

クラウド・テクノロジー活用部会の活動の一環として、[気象庁より試行配信中の気象庁XML](#)を受信し、簡単に参照できるAPIを作成してみました。

- ・蓄積データ参照&REST API
- ・SPARQL クエリ発行
- ・WebSocketによる配信

蓄積データ参照 & REST API

- 簡単に参照でき、検索機能も豊富
- 一番、アクセス数が多い、search → XML取得



気象庁防災情報 XML 一覧

注意事項 2019-09-17 「蓄積データ参照&REST API」では大雨危険度通知を利用できません。「SPARQLクエリ発行」「WebSocketによる配信」を使用してください。

変更履歴 2014-06-26 API の応答に ISO8601 形式の日時が追加されました。

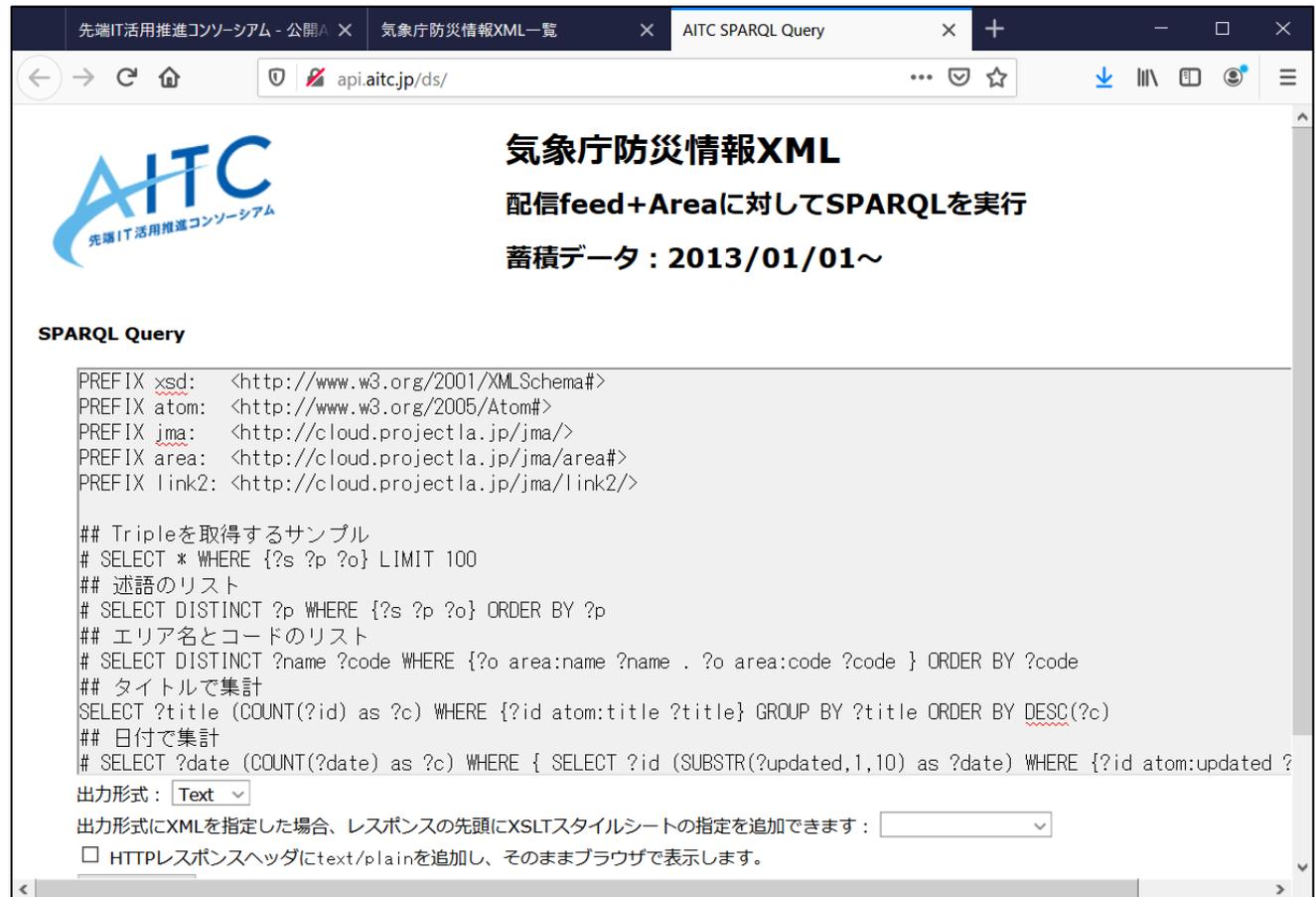
最近の配信：

received	datetime	status	infotype	editorialoffice	title	eventid
2020-09-01 10:05:45.26782	2020-09-01 10:04:47.0	通常	発表	神戸地方気象台	気象特別警報・警報・注意報	-
2020-09-01 10:05:45.167433	2020-09-01 10:04:47.0	通常	発表	神戸地方気象台	気象警報・注意報	-
2020-09-01 10:05:45.013146	2020-09-01 10:04:47.0	通常	発表	神戸地方気象台	気象警報・注意報 (H27)	-
2020-09-01 10:01:23.33309	2020-09-01 10:00:56.0	通常	発表	名古屋地方気象台	生物季節観測	20200901100000_57

一番最後に
PUSH配信された
気象庁XML

SPARQLクエリ発行

- クエリを書けば、サーバ側で集計処理
- SPARQLを発行できるサーバ自体、珍しい



先端IT活用推進コンソーシアム - 公開 / 気象庁防災情報XML一覧 / AITC SPARQL Query

api.aitc.jp/ds/

気象庁防災情報XML
配信feed+Areaに対してSPARQLを実行
蓄積データ：2013/01/01～

SPARQL Query

```

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX atom: <http://www.w3.org/2005/Atom#>
PREFIX jma: <http://cloud.projectla.jp/jma/>
PREFIX area: <http://cloud.projectla.jp/jma/area#>
PREFIX link2: <http://cloud.projectla.jp/jma/link2/>

## Tripleを取得するサンプル
# SELECT * WHERE {?s ?p ?o} LIMIT 100
## 述語のリスト
# SELECT DISTINCT ?p WHERE {?s ?p ?o} ORDER BY ?p
## エリア名とコードのリスト
# SELECT DISTINCT ?name ?code WHERE {?o area:name ?name . ?o area:code ?code } ORDER BY ?code
## タイトルで集計
SELECT ?title (COUNT(?id) as ?c) WHERE {?id atom:title ?title} GROUP BY ?title ORDER BY DESC(?c)
## 日付で集計
# SELECT ?date (COUNT(?date) as ?c) WHERE { SELECT ?id (SUBSTR(?updated,1,10) as ?date) WHERE {?id atom:updated ?

```

出力形式:

出力形式にXMLを指定した場合、レスポンスの先頭にXSLTスタイルシートの指定を追加できます:

HTTPレスポンスヘッダにtext/plainを追加し、そのままブラウザで表示します。

WebSocket & Comet

- サーバ無しでも、リアルタイムに受信できる
- デモサイトで使用されていた



The screenshot shows a web browser window with the following content:

- Browser Tabs:** 先端IT活用推進コンソーシアム, 気象庁防災情報XML一覧, AITC SPARQL Query, WebSocket
- Address Bar:** api.aipc.jp/websocket/index.html
- Page Title:** 気象庁防災情報XML受信ツール
- Section:** WebSocket版
- Text:**

WebSocket+Kafkaにより、サーバ側で受信した防災情報XMLを配信します。

注意：Proxyを経由すると、WebSocketが接続できない場合があります。

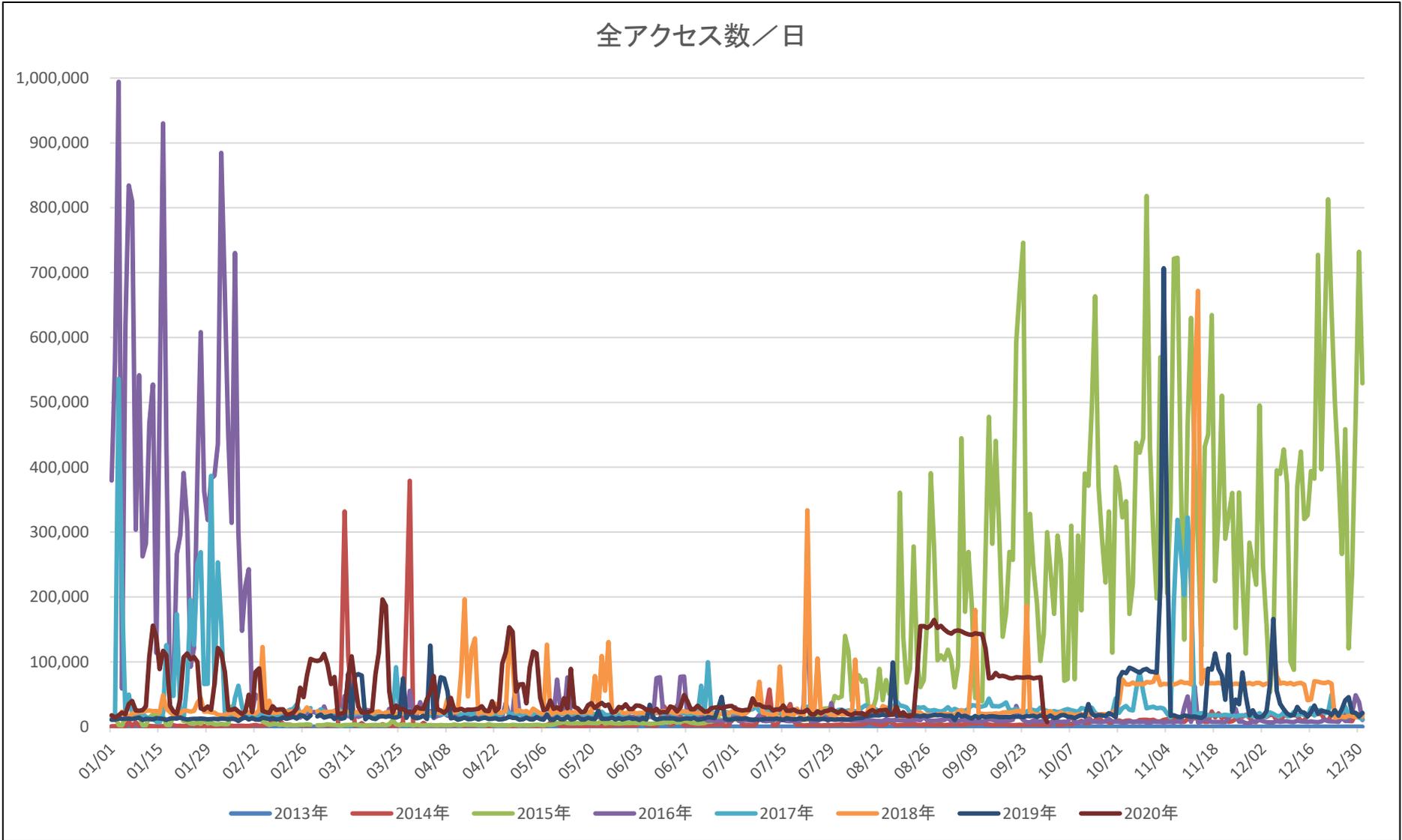
注意：50秒間隔でハートビートを行います。

注意：受信した防災情報XMLをすべてブラウザのメモリ上に貯めるので、定期的に再接続してリフレッシュしてください。

[Comet版はこちら](#)
- Controls:**
 - Buttons: 再接続, 切断
 - Input field: 気象警報・注意報
 - Buttons: filter, xpath
 - Input field: //jmx:Title[text()='気象警報・注意報']

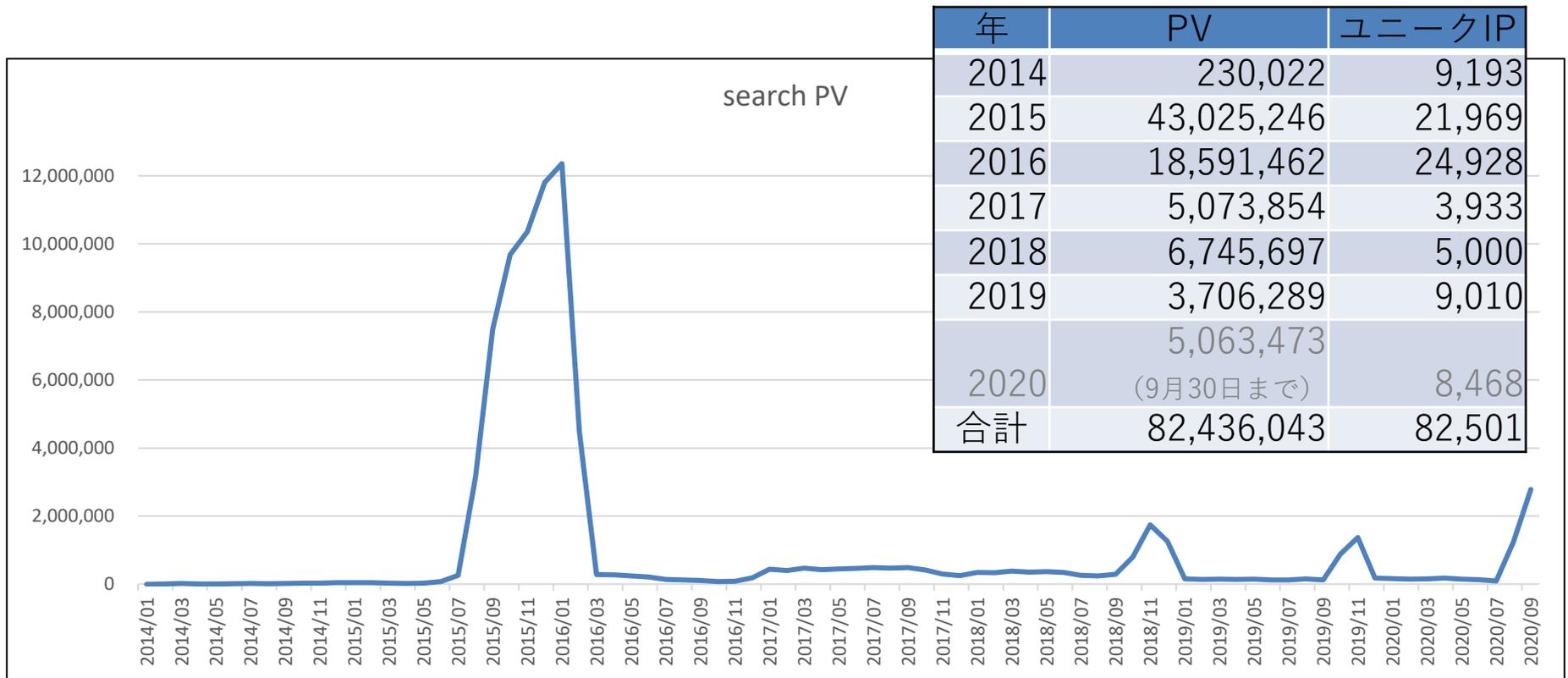
アクセス数／日：全リクエスト

全アクセス数／日



アクセス数／月：REST

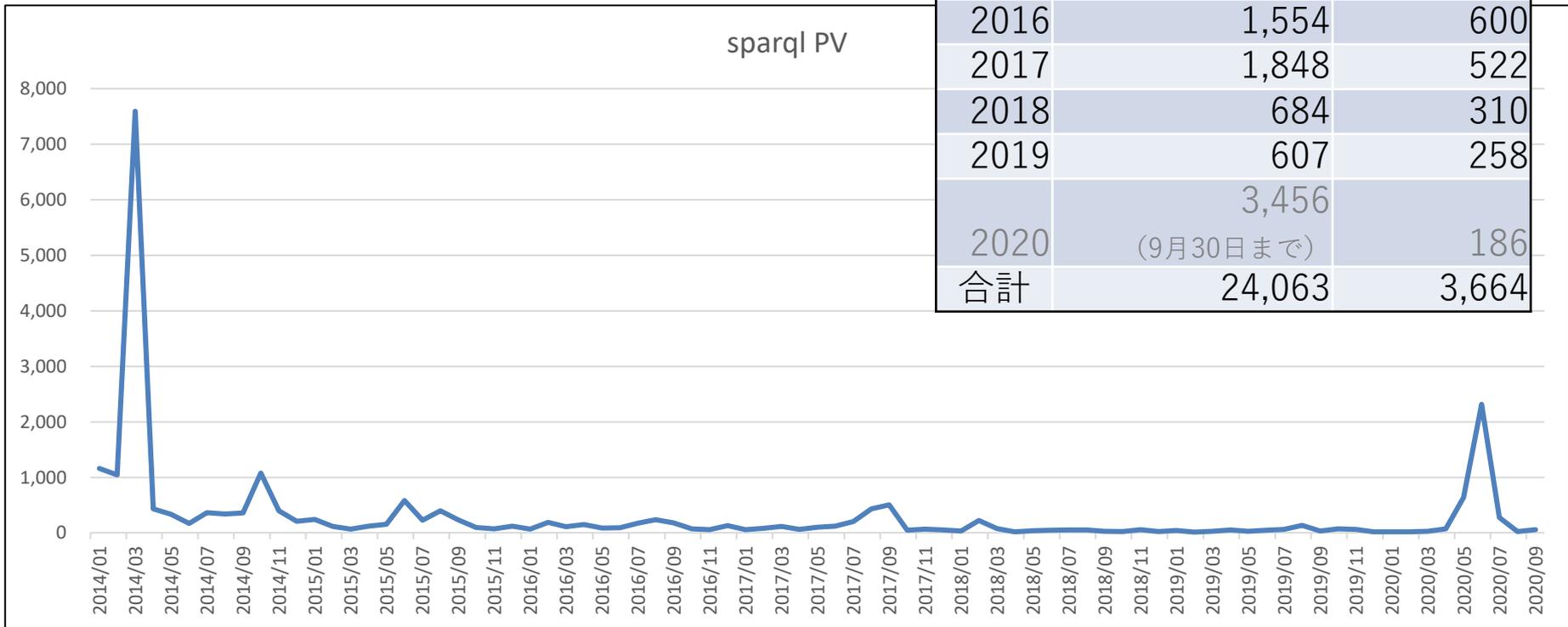
- searchリクエストのみ
- 毎年11月前後にアクセス数が増えるのは、卒業研究？
- 2015年は異常にアクセス数が多かったけど、問題なし



アクセス数／月：SPARQL

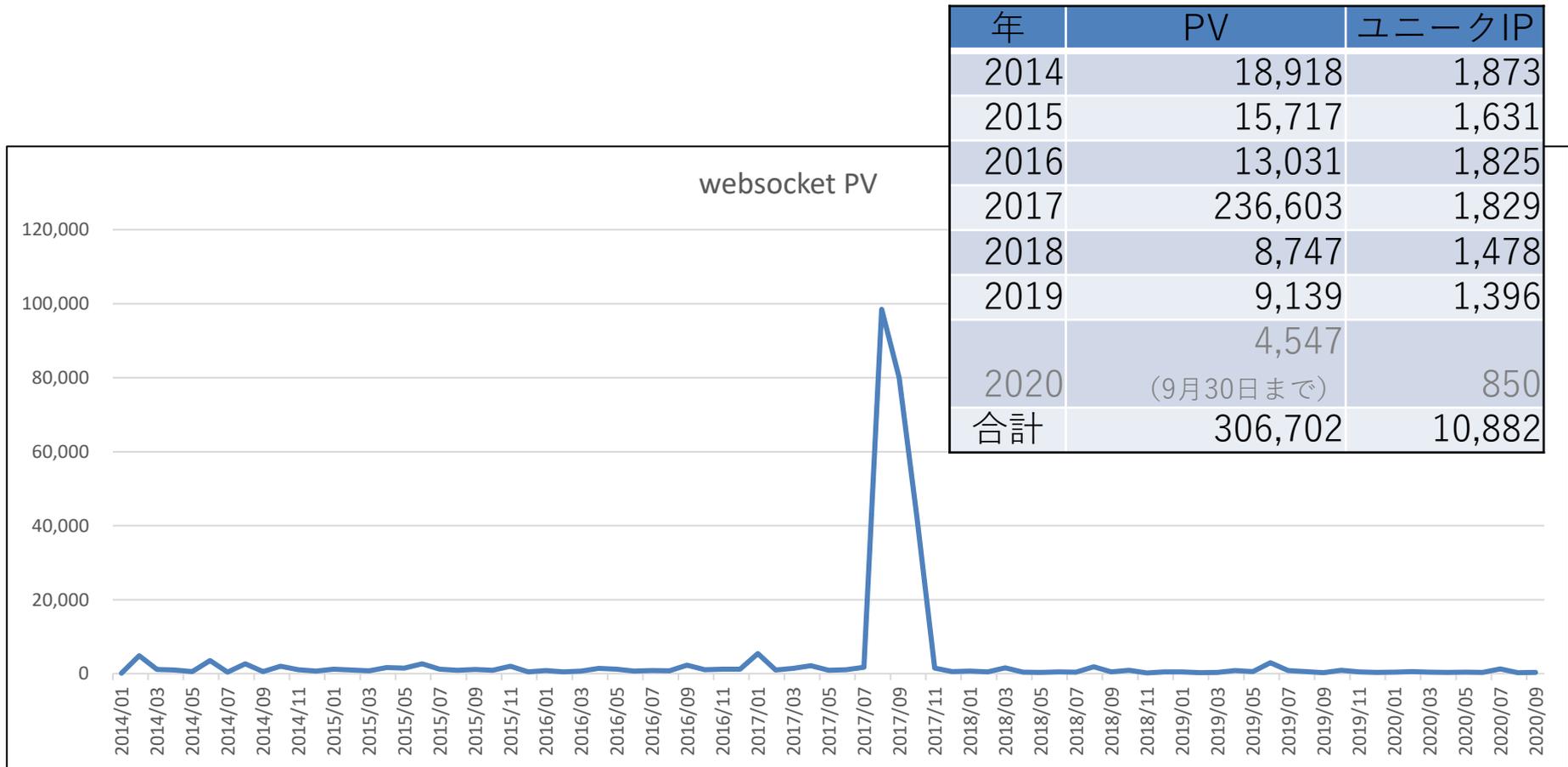
- 2014年頭は、福岡工業大学の卒業研究のアクセス
- 2020年6月は、不明

年	PV	ユニークIP
2014	13,484	922
2015	2,430	866
2016	1,554	600
2017	1,848	522
2018	684	310
2019	607	258
2020	3,456 (9月30日まで)	186
合計	24,063	3,664

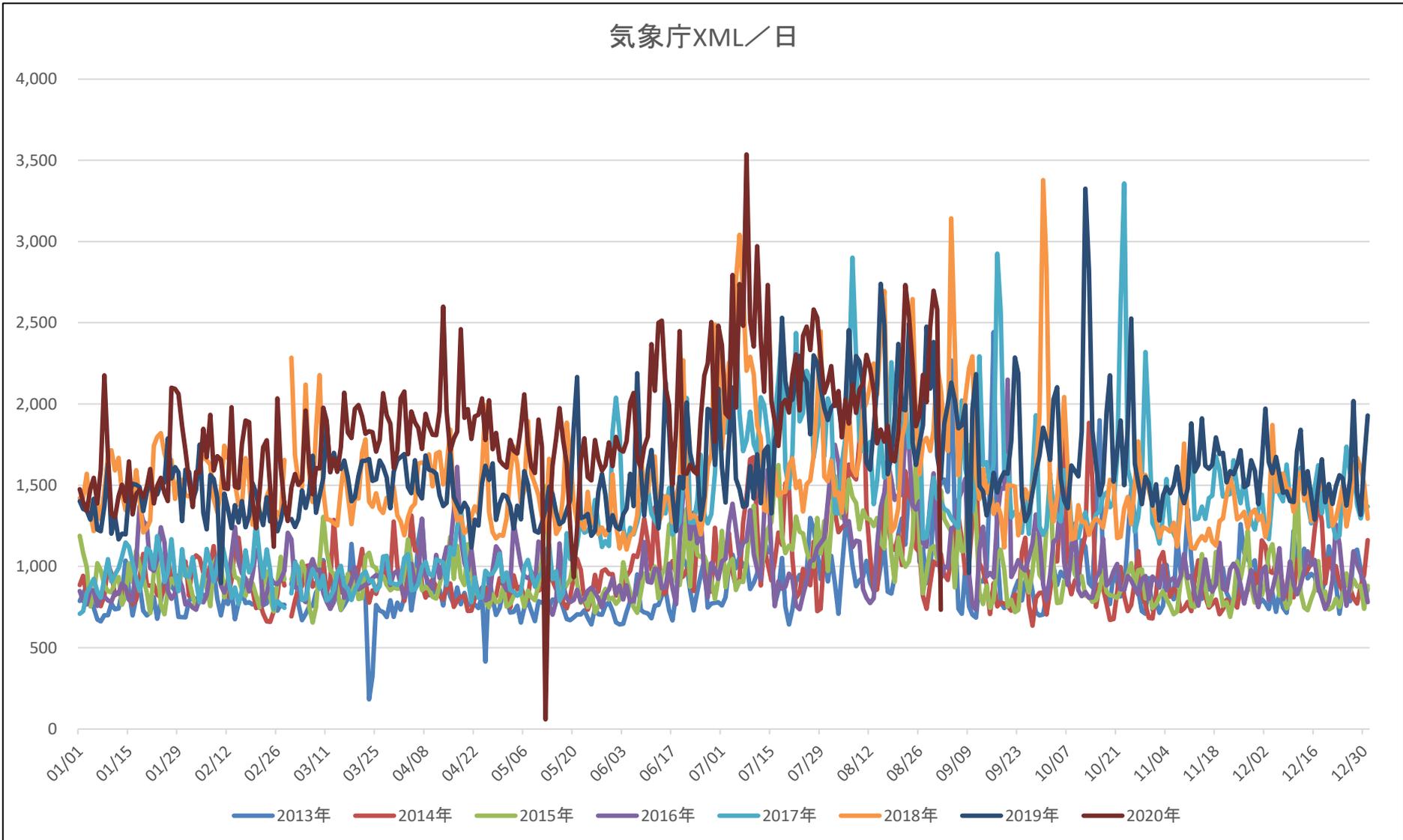


アクセス数／月：Websocket

- 不定期に、同じIPから大量の同時接続がある



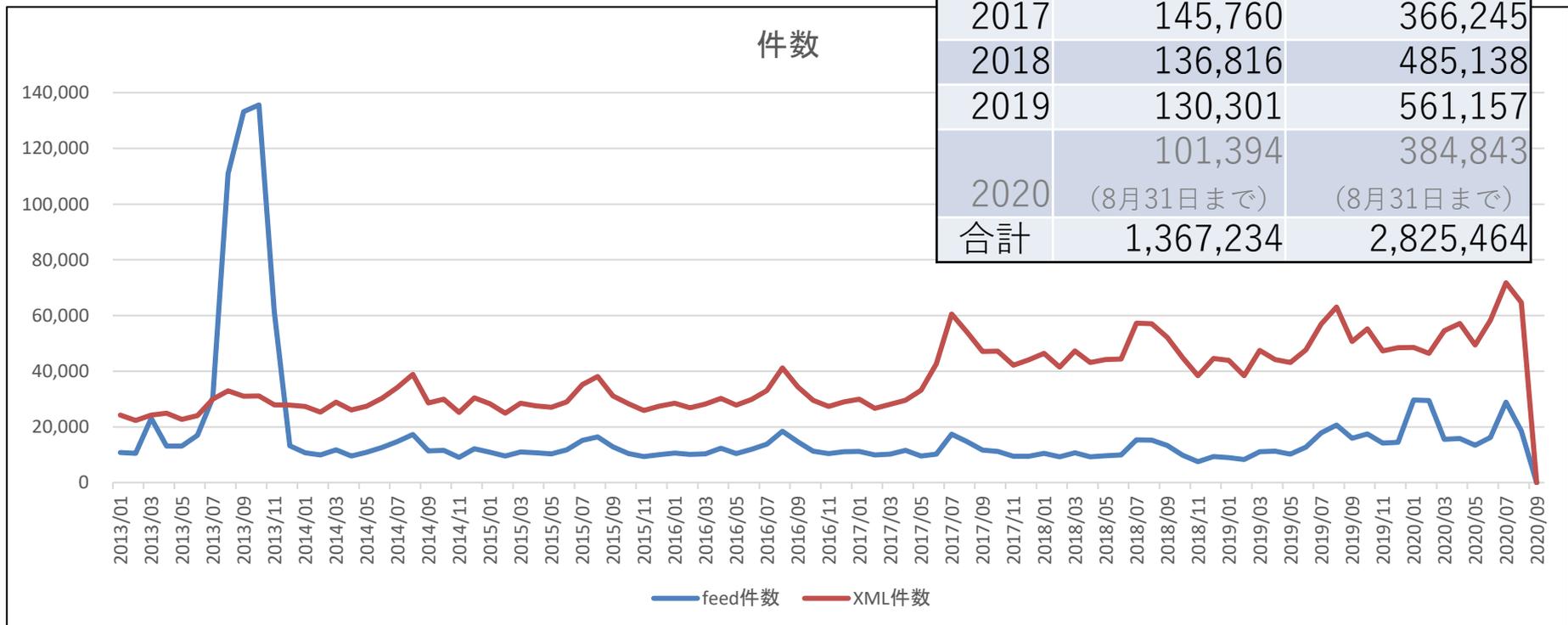
気象庁XML配信件数／日



気象庁XML配信件数／月

- 2013年9月頃には、同じfeedが40通届いていた
- Google Hubが配信モレを防ぐために、重複して配信？

年	feed件数	XML件数
2014	572,597	323,554
2015	141,832	352,903
2016	138,534	351,624
2017	145,760	366,245
2018	136,816	485,138
2019	130,301	561,157
2020	101,394 (8月31日まで)	384,843 (8月31日まで)
合計	1,367,234	2,825,464



気象庁XML配信サイズ／月

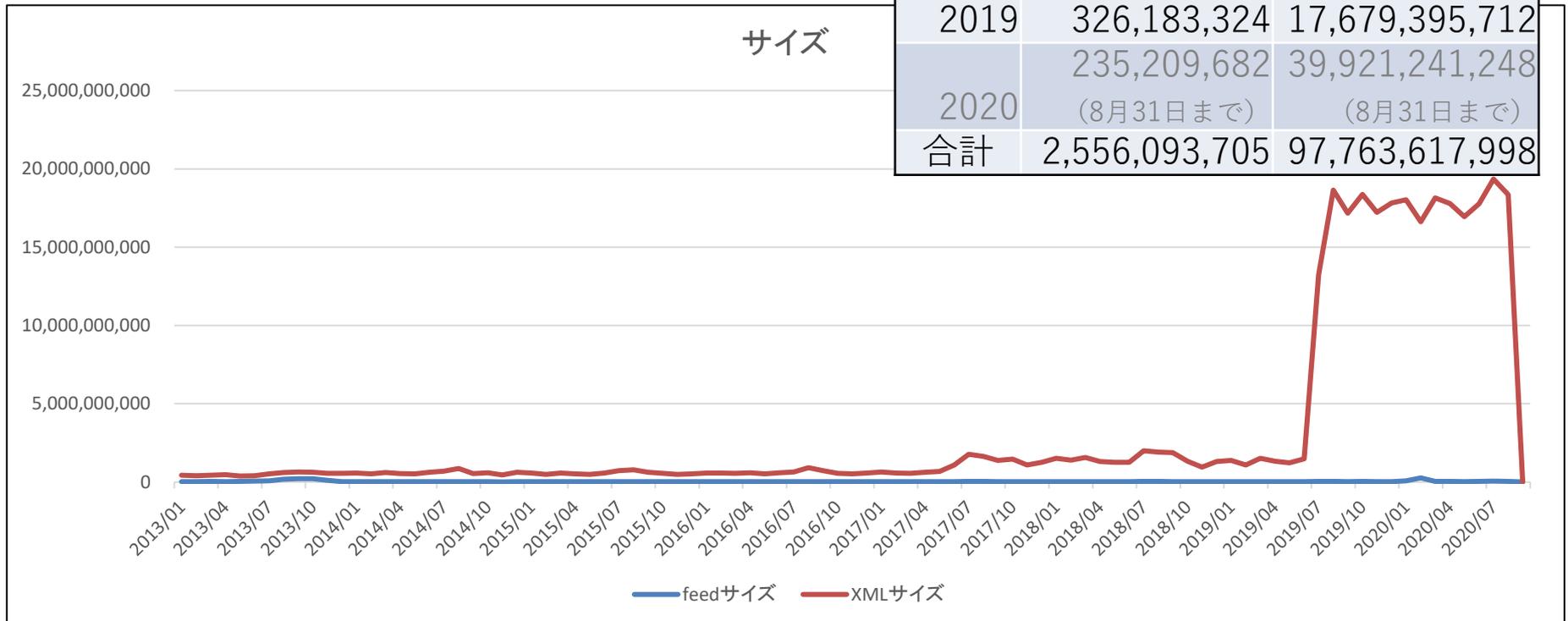
- 2019年7月から、総XML配信サイズが10倍になった

2019/7/10「大雨危険度通知」配信開始

1XMLのサイズ:3.7MByte

10分周期で3.7MByteが配信

年	feedサイズ	XMLサイズ
2014	1,020,050,329	6,015,465,297
2015	227,187,786	7,150,183,301
2016	222,141,436	6,914,619,663
2017	235,307,719	7,327,332,458
2018	290,013,429	12,755,380,319
2019	326,183,324	17,679,395,712
2020	235,209,682	39,921,241,248
	(8月31日まで)	(8月31日まで)
合計	2,556,093,705	97,763,617,998



外部からの問い合わせー1

- 2014年10月 福岡工業大学
 - 以前のハンズオンのSPARQLがタイムアウトする
 - 卒業研究で学部の学生に利用させる予定
- 2015年03月 日本ア〇・〇ー・〇〇
 - 地震発生時の弊社拠点への影響把握の一環として、気象庁XMLデータをAPI経由で参照している
 - API公開をもっと長く続けて欲しい(2016年8月終了予定だった)
- 2016年05月 株式会社〇〇〇〇っと
 - スマートウォッチのウィジェットで、APIを利用したい
- 2016年06月 株式会社〇ーガ
 - GoogleMaps上に、災害情報を可視化するデモサイトに使用
 - API公開をもっと長く続けて欲しい(2016年8月終了予定だった)

- 2017年06月 香港(英文)
 - APIの使い方を教えて欲しい
- 2017年07月 個人
 - 自身のサイトでWebsocketを利用している
 - WebsocketのAPIが利用できなくなっていると連絡以降、止まるたびに連絡がきた
- 2017年07月 学生
 - 大学のセミナーでapiを使った作品を作る
 - apiキーのようなものはあるのでしょうか？
- 2020年05月 奈良先端大学 学生
 - 気象庁防災情報XMLを用いた研究を進めたく、REST APIを利用
 - 試行錯誤している間にInternal Server Errorが出てしまった

その2へ

トラブルと解決方法: SPARQL

- クエリを何度も実行すると、メモリ不足でタイムアウト
 - Javaとミドルウェアをバージョンアップし、メモリ割り当てを追加
 - 実験環境で練習してから、本番環境で実施
 - データは別形式で保存しておいたので、再登録を実施
- 一気に配信がくると、追加処理が間に合わない
 - SPARQLの追加処理だけ、1日1回のバッチ処理にした

- WebSocketのプロトコルが一部のProxyを通過できない
 - 443ポートを使って、たいていのProxyを通過できるようにした
- 一定期間周期で、参照できなくなる
 - 内部的に Int で管理しており、数か月で int の範囲を超える
 - 暫定対応: 定期的のリセット
 - 恒久対応: ミドルウェアのバージョンアップ
 - Byte単位 → 件数単位になったことで、int の範囲を超えない
- 不定期に、同じIPから大量同時接続
 - 海外からが多かったので、海外からは iptables で接続拒否

- 気象庁XMLを使いたい人
 - 気象庁サイトのPULL型を利用
 - 簡単に参照はできるけど、履歴は1週間分のみ
 - 気象業務支援センターを利用
 - 負担金が必要
 - 他の蓄積サービスを利用
- AITCの気象庁XML取得APIの今後
 - 2021年2月末まで、そのまま置いておく予定
 - データは2020年8月末までのものしかない
 - 蓄積された大量データを使って部会内で実験
 - 気象庁XML: 約282万件
 - アクセスログ: 約8,000万件

AITC クラウド活用部会への参加、
お待ちしております



<http://aitc.jp>



<https://www.facebook.com/aitc.jp>



ハルミン

AITC非公式イメージキャラクター