

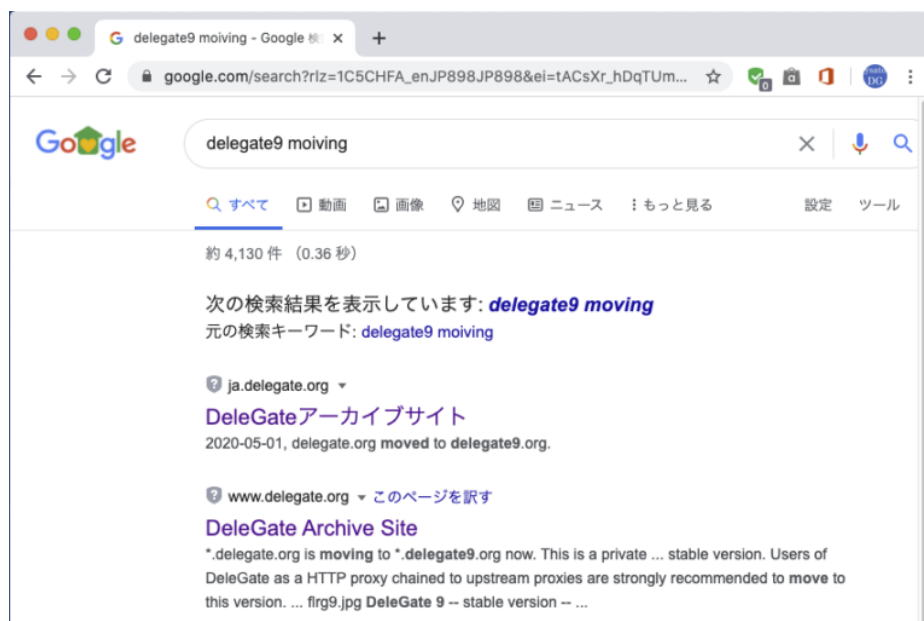
株式会社 ITS MORE

2020年4月始動

2020年5月1日 投稿者: YSATO@DELEGATE.ORG

本日引っ越し決行

無駄な出費を一刻も早く止めるべく、delegate.org の旧サーバを停止しました。新ドメイン delegate9.org への引っ越し案内が Google と Yahoo の検索でヒットするようになったので、もういいかなと。Google のキャッシュでは以下のように表示されます。よい卒業記念写真になりました。



DeleGate Archive Site (www.delegate.org)

Official Site | Updates | Download | Manual | Documents | Forum | Search | Stats

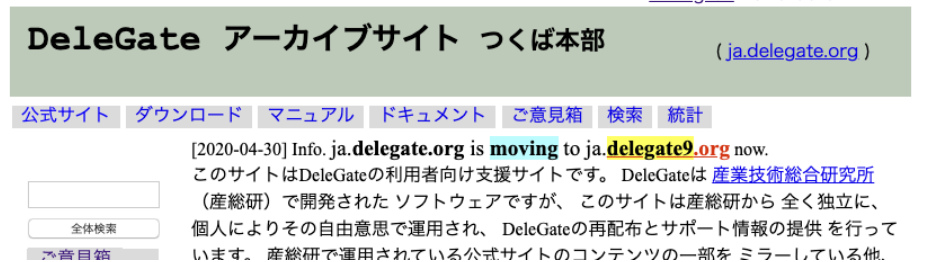
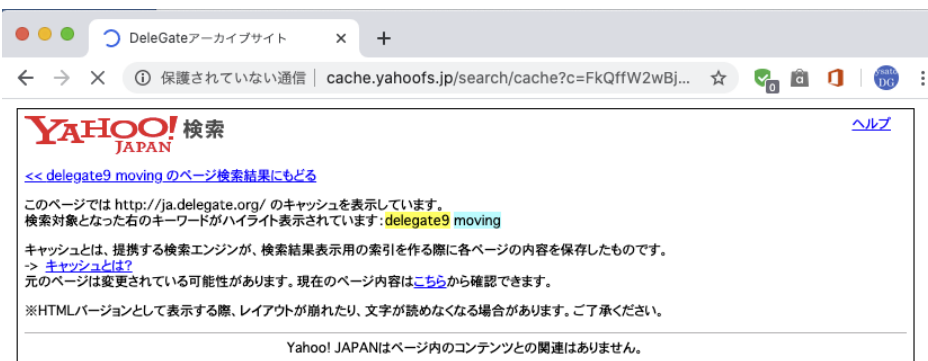
[2020-04-30] Info. ***.delegate.org** is moving to ***.delegat9.org** now.
 This is a private and voluntary site that is independent of **AIST**, for providing support information and redistribution of DeleGate, a general purpose proxy server software developed in AIST and, more specifically, for providing voluntary-based support for users in forums, supplementary contents for guide, some experimental services and demonstrations, as well as contents mirroring some of those in the official site of DeleGate of AIST.
 So, to get original distribution files, documents, license information, and so on, please visit the [official site](#). 日本語

DeleGate アーカイブサイト つくば本部 (ja.delegate.org)

公式サイト | ダウンロード | マニュアル | ドキュメント | ご意見箱 | 検索 | 統計

[2020-04-30] Info. **ja.delegate.org** is moving to **ja.delegat9.org** now.
 このサイトはDeleGateの利用者向け支援サイトです。DeleGateは [産業技術総合研究所](#) (産総研) で開発されたソフトウェアですが、このサイトは産総研から全く独立に、個人によりその自由意思で運用され、DeleGateの再配布とサポート情報の提供を行っています。産総研で運用されている公式サイトの内容の一部をミラーしている他、独自のコンテンツとして、フォーラムにおけるボランティアベースの利用者支援、ガイドのための補完文書、いくつかの実験的サービスやデモを提供します。
 オリジナルのコンテンツ (全配布ファイル、文書の原本、ライセンス情報、など) は [公式サイト](#) のほうで閲覧・取得してください。

すでにかなり昔からGoogleしか使わないので、Yahooでもキャッシュのサービスをしているというのは知りませんでした。こんな感じでした↓。まがまがしく「キャッシュ」と書いているのは対抗心でしょうか(笑)。いつキャッシュしたのか表示しないのは、何かの配慮なのでしょう吗？



Googleのキャッシュではカエルのインラインイメージが表示されていますが、これ

はただ私のブラウザが、まだそれをキャッシュしているためです。ではなぜYahooのほうではそれが表示されず壊れているのでしょうか？両者ともキャッシュの説明に、

「<http://www.delegate.org/>のキャッシュです」と言っていますが、実はそのURLは <http://www.delegate.org/delegate/> へのリダイレクションになっています。その `index.shtml` の中のソースにはカエルのインラインイメージをこう書いています。

```
IMG ALT=DeleGateIcon SRC=DeleGateLogoTrans.gif
```

Googleではこれがきちんと、`/delegate/DeleGateLogo...` となりますが、Yahooでは `/DeleGateLogo...` になってしまっているわけです。そんなリソースはありませんから、表示できなくて壊れる。つまり、Yahooのキャッシュは、`/` が `/delegate/` にリダイレクション (302 Moved) される前のベースURLを使ってしまっている。バグですよ。

まあ、こういうこともあるので、小さなインラインイメージは `data:img` URLで埋め込みたいところです。

なんで後追いなのに、こういう基本的なところで先駆者に劣るのか、わかりません。キャッシュ日付の件だけでなく、Googleでは案内の中に「フルバージョン | テキストのみのバージョン | ソースを表示」とリンクがあり、簡潔にして完備しています。こういう機能を使いそうなユーザ層を考えれば妥当な機能提供だと思います。

ところで、こういう魚拓は感傷的な記念写真に留まらず、「その時にはその文書が存在した」という証明として有益です。ただし、Googleはそれに責任も持たないでしょうし、キャッシュ保存期間もあまり長くないようです。(3月末まで存在した旧「公式サイト」delegate.hpcc.jpのGoogleキャッシュも、今はほぼ消えました。)

まあ、プライバシーの問題とかで速く消したい人のほうが多いでしょう。でも、どこだったか忘れましたが、海外ですごい古い版の魚拓まで保持しているサイトを見たことがありますから、利用できるかも知れません。まじめに存在証明をしたかったら、タイムスタンプ付き電子署名をしなくてははいけません。電子署名付きでどこかに永久保存されてれば完璧ですね。まあ、証明できるのは、ベースになる静的テキスト部分だけでしょうけど。

それで気になったので、Googleの画像検索を見に行ったのですが、やはり画像の「キャッシュ」機能は無いようでした。

ウェブロボットがHTTPのリダイレクションを追うか否かは、ロボットの気性や目的

を表していると思います。私自身も全文検索エンジン用のロボットを作った時に、「そのサイト以外まで追うか」というところで少し考えましたが、どこまで追うかという段数を指定できるようにしました。これは、その文書と関連するページをあぶりだすシンプルな方法だと思っています。

それで、今回引っ越しの際に数時間行った移転先へのリダイレクションを、ロボットが追うかどうか、観察すると興味深い状況が見られました。「ドメイン名が逆引きできないところから来るロボットはリダイレクションを追わない」というものです。そういった素性の知れないロボットの来訪目的は、クロールした結果を検索サイトに載せるとかではなく、ターゲットとしたサイトの中身を探るといような事なんだろうなと思いました。

今回の引っ越しのおかげで、出来の悪い荒らしロボットはしばらく来ないと期待されます。おかげで、大変静かになり（笑）、アクセスログを見る気になりました。久しぶりに `sed 's/.*//'| sort | uniq -c | sort` です。原始的ですねえ。すると、大変懐かしいロボットさんたちばかり来てました。googlebot.com, msn.com, yandex.com、おひさしぶりでした。ちっ、your-server.de も来てやがる… 一方、ほとんど人間は来てないようでした（笑）

やくざなロボットには来て欲しくないですが、立派なロボットさんは歓迎です。主に検索サイトのロボットですが、もはや検索サイトのシェアはGoogleとYahooを合わせると90%以上、日本に限れば99%以上になっているそうです[[日本・世界の検索エンジンシェア](#)]。そういう意味では、もうこの両社、とMSN以外のロボットはお断りしても良い。

人間様に迷惑をかけずにロボットだけを撃退しようと、過去、いろんな工夫をしましたが、結局そのためにかなりの労力も、コンピュータのコストもかかりました。人間様がめったに来なくなった今としては、優良ロボットだけを顔パスでお通しして、あとは「私はロボットではありません」ボタンで撃退すれば良いかなと思います。

アドレス出自で優良かどうかをスクリーニングする（ホワイト国って言いましたっけ）のは簡単ですが、もしまじめな無名ロボットさんが来た時に門前払いして良いものか。どうやら、振る舞いを追跡観察することで、多少そのあたりを推定できるような気がしています。

あとは、この検問のためにアプリケーションレベルで判定していると、結局コンピュータに負担をかけ、それなりの費用がかかってしまいます。なにせ従量課金なん

で。とすれば、ネットワークインタフェースレベルで、IPアドレスレベルで塞いでしまうことです。できればドメイン名でも塞ぎたいところですが、IPルータとはレイヤが違うので無理。

こいつウザいから門前払いしよう、という判断はDeleGateでもやっています。その結果（IPアドレスとマスク）をルータなりに、機械的に設定すれば良いわけです。Azureでそれをやる場合、「ネットワークインタフェース」というリソースの設定用APIをAzureが公開しているかということが問題になりますが、まあ人間もウェブブラウザからやってるんで、HTTPでできることは確かです。あとは料金。まさか、フィルターのテーブルサイズに従量課金とかしてないよねAzure…

とは言え、そういう明らかに必要で金になりそうなところをMSが放っておくはずはありません。きっと、それに相当する頭のよいリソースを提供している事でしょう。従量課金で X-D

2020-0501 sato@izmoh