

株式会社 ITS MORE

2020年4月設立

2020年5月27日 投稿者: YSATO@DELEGATE.ORG

私はロボットです

社長：昔 robots.txt を無視するロボットに業を煮やして、私はロボットです (ImRobot) 、ロボットではありません (ImNotRobot) というクッキーを作りましてね。ヘビーにアクセスされると困るリソースから排除を試みたりしてました。

開発：その後ロボット対策に、ぶにゃぶにゃの文字とか画像とかを人間に認識させる対策が流行りましたが、あれってちょっと人間にはつらい仕打ちでした。ああいう画像認識はロボットのほうがよっぽど得意なんじゃかと思うんですが。

社長：だから、Google reCAPTCHA みたいのが現れた時は、救世主かと思いましたよ。あれを Google からまったく切り離れた中立的なサービスなり製品として提供してくれると良いのですが。reCAPTCHA バッジの中のプライバシーとか利用規約だのをクリックして読むと、一体なんの事かと恐ろしい感じがします。

開発：Google reCAPTCHA を Google のロボットが突破できるのか、興味ありますね。盾鉾試験はしてるんでしょう。

社長：昔、自前で全文検索エンジンを作ってたんですが、主な検索対象はローカルなファイルでした。そこそこ気に入ってたし、今でも DeleGateミュージアムで動いています。ですが、Google から [Google Desktop](#) サーチが出て、ああもういらなくなかって、開発をやめてしまいました。それが Google Desktop が 2011年に撤収してしまって。まあ reCAPTCHA はやめないだろうけど。

開発：Mac では Spotlight のごり押しがひどくてみんなから嫌われてましたが、最近使ってみたら面白いなとは思いました。

社長：でも結局私が欲しいのは、grep なんですよね。正規表現で検索できなくちゃ。そりゃあ Google みたいな巨大なデータには適用できないだろうけど、自分のローカルにある検索したいファイルの中のテキスト部分なんて、高々100MB～1GB程度なわけです。全然 grep で行けます。1秒で grep できちゃうし、vi でだって編集できる。私は、find もかったるいので、tar でまとめて vi で検索なんてことをよくやります。

開発：コンピュータのプログラムに関しては、Eclipse とかの IDE の検索機能がとても良いですね。やはり形式が規定されたデータの間の参照関係は役に立ちます。いっそ文書も全部、Eclipseで検索してやろうかと思いましたが、さすがにうまく行きませんでした。

社長：でも結局、IDEで使ってたのも、主には grep 的なテキスト検索でしたね（笑）。正規表現でマッチして置き換える。これ、半世紀近く前 sed の時代に完成していて、今は退化したか社会の片隅化してしまった技術の一つだと思います。

開発：コンピュータのパーソナル化で「直感的に操作できること」に重心が移りすぎたですかね。正規表現だなんて抽象的なものは、火炙りの刑というか。

社長：たとえば干物みたいな、流通技術が乏しかった時代の対処技術は、実はものを美味しくする技術でもあった、みたいなところもありますよね。

開発：最近少しだけ仕事した FPGA の統合開発環境もとても印象的でした。ハードウェア部品とは言え、結局は文字列か数値で表現されたデータのマッチングではあるのですが、図的な表示との双方向の検索は、ちょっと感動ものというか。

開発：まったく昔に戻る事なんてあり得ないけど、昔の本質的な技術とのベストミックスを考えると良さそうにも思います。

開発：タイトルと全然違う話になりました。

社長：まあ検索ということで、なんとか bot 繋がり。

—

2020-0527 SatoxITS

