

Startup of PeWeProxy

Michal Barla, Tomáš Kramár, Jozef Tomek

<http://peweproxy.fiit.stuba.sk>

Otázky? :)

Motivácia

- Pomôžete pri vyhodnocovaní záv. projektov
- Prispejete k vytvoreniu datasetu, ktorý môžete neskôr použiť aj vy
 - myslite na svoje DP, DizpP :)
- Bude sa vám na webe surfovať lepšie
 - rozšírenie google dopytu
 - vylepšenie navigácie na fakultnej stránke
 - poznámkovanie textu
 - sociálne odporúčania

Čo treba spraviť?

- Nastaviť surfovanie cez proxy server
 - PAC súbor
- Nastaviť si identifikátor v user-agent
 - firefox extension
 - java browser patcher (FX, Opera, IE, Chrome)
 - ručne
- (ideálne) zabudnúť na proxy a surfovať ako keby nič
 - FoxyProxy
- Vo všetkom pomôže stránka peweproxy.fiit.stuba.sk

Čo sa loguje?

- Zaznamenávame
 - URL
 - identifikátor (nespojený s menom)
 - časovú pečiatku
- Nezaznamenávame
 - request data
 - response data – s výnimkou extrahovaných kľúčových slov
- Cez proxy vôbec nejde https://

Obavy?

- Pozrite si vaše logy (len vaše logy)
- Vymažte to, čo tam nechcete mať
 - a naozaj sa to vymaže :)
- Preštudujte si zdrojáky, sú na webe
 - Java
 - Ruby on rails
- Všetko cez <http://peweproxy.fiit.stuba.sk>

Čo sa chystá?

- Plán je 18.4. nasadiť query disambiguation
- Intenzívne sa pracuje na vrstve nad fiit.stuba.sk
- Detekcia virtuálnych komunít
- Ak chcete byť informovaní
 - RSS kanál
 - notifikácia e-mailom

Máte nápad?

- Jednoduchá možnosť obohatiť proxy o novú funkcionálnosť cez plugíny
 - odchytyvanie, modifikácia a presmerovávanie requestov
 - modifikácia response (String, DOM)
 - injektovanie javascriptu a odchytyvanie jeho volaní

Posuňte to ďalej

- Budeme radi, ak na to namotáte ďalších ľudí mimo našu komunitu
 - chceme rozmanité dáta, v ktorých sa dá dolovať, zhlukovať, klasifikovať...

<http://peweproxy.fiit.stuba.sk>