

Posudek oponenta diplomové práce

Autor/autorka práce: Jindřich Pouba

Název práce: Klasifikace uživatelů přistupujícím k webovým službám

Obsah práce

Cílem práce je navrhnout a implementovat systém, který umožní klasifikovat uživatele webových služeb na základě jejich chování. V úvodních teoretických kapitolách jsou představeny jednotlivé technologie a postupy, které jsou následně použity v praktické části. Bohužel tato část práce by si zasloužila výrazně lepší zpracování. Některé popisy, především fungování webových technologií, jsou velmi povrchní a mnohdy nepřesné. Student například staví na stejnou úroveň technologie VPN a Tor, kde jejich použití je zcela odlišné. Stejně tak některé závěry, které jsou z teoretické části vyvozeny a následně použity v praktické části, jsou značně nepřesné. Především možnosti rozlišení jednotlivých uživatelů, kteří k webovým službám přistupují prostřednictvím NAT, proxy nebo VPN. U fungování webových služeb zcela postrádám možnosti přesnější identifikace uživatelů například pomocí sessionID.

V praktické části práce student popisuje existující metody klasifikace uživatele. Tato klasifikace je téměř vždy omezena pouze na rozpoznání lidského uživatele a robota, což je škoda, neboť možnosti klasifikace, především z pohledu marketingu, mohou být výrazně širší a mohou přinášet vyšší přidanou hodnotu. Na závěr praktické části jsou zdokumentována podrobná měření existujících i studentem nově navržených postupů.

V závěru práce jsou krátce shrnuty jednotlivé měření. Ani v závěrečné kapitole není uvedeno čím bylo měření provedeno, teprve v přílohách se čtenář dozví, že student vytvořil program, kterým jednotlivé metody testuje, což je s ohledem na odvedenou práci nešťastné.

Kvalita řešení a dosažených výsledků

Zásadním nedostatkem práce je její teoretická část, ve které jsou nepřesnosti a celkově by si zasloužila více péče. Stejně tak by bylo vhodné věnovat více péče zdrojovým kódům, které téměř neobsahují komentáře a ty které v kódu jsou, což je většinou jediný komentář na jeden soubor, jsou spíše zavádějící. S ohledem na fakt, že je program napsaný v programovacím jazyce Java, považoval bych za vhodné program doplnit o Javadoc. Zároveň by bylo vhodné na přiloženém CD uvést alespoň základní readme, které umožní lepší orientaci v adresářích. Pokud student vytvořil nový testovací nástroj, jak se dozvídáme v přílohách, bylo by vhodné v práci představit existující řešení a uvést proč bylo nutné vytvořit nástroj nový. Co je provedeno velice detailně a kvalitně je popis jednotlivých metod rozpoznání člověk a robota a následná měření. V této části práce je vidět, že student nastudoval mnoho teorie a provedl velké množství řešení, které detailně zdokumentoval.

Formální úroveň

Rozsah práce se svými více jak 100 stranami textu je nadstandardní. Bohužel se v textu vyskytuje nemalé množství chyb a překlepů. Struktura práce je přehledná a logicky členěná. Jako nevhodné

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky

vidím především přesunutí všech informací o vytvořeném programu až do části s přílohami. Zároveň v závěru práce bych uvítal shrnutí cílů, představení odvedené práce a zhodnocení výsledků. Bohužel závěr se omezuje pouze na hodnocení jednotlivých použitých metod klasifikace. Ani v závěru není nijak zněněno, že student vytvořil vlastní testovací nástroj na jednotlivé klasifikace. V práci není obsažen ani seznam obrázků či tabulek. Stejně tak postrádám seznam zkratk.

Práce s literaturou

Práci s literaturou mi nepřijde zcela ideální. Práce sice obsahuje 27 odkazů na literaturu, ale v 25 případech se jedná o odkaz na webové zdroje a ve dvou případech dokonce je zdrojem Wikipedie, což není jako vhodný zdroj pro tuto práci.

Splnění zadání

Zadání bylo splněno.

Dotazy k práci

1. Proč je k identifikaci jednotlivých návštěvníků použita pouze IP adresa, když tento údaj mnohé technologie, jako NAT, proxy či VPN, překrývají? Nebylo by možné jednotlivé návštěvníky identifikovat výrazně přesněji pomocí dalších údajů jako je sessionID, IP a zároveň PORT klienta, HTTP autentizace či VPN IP?
2. Popsané metody se defakto vždy omezují na klasifikaci zda se jedná o člověka či robota. Je možné v rámci zkoumaných metod identifikovat i další chování uživatelů jako například zda se jedná o náhodného či stáleho zákazníka?
3. S jakými již existujícími nástroji na klasifikaci uživatelů webových služeb jste se seznámil, než jste začal vytvářet nástroj vlastní a v čem byly tyto nástroje nedostatečné?

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 6.6.2017

Ing. Luboš Matějka

**SOUHLASÍ
S ORIGINÁLEM**

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky

①