

Posudek bakalářské práce

Autor: Tomáš Vyskočil

Název: Detekce anomálií v chování uživatelů

Posudek vypracoval **vedoucí práce:** Ing. Jan Šedivý, CSc

Tomáš Vyskočil se ve své bakalářské práci zabýval aktuálním problémem detekce anomálií v chování uživatelů, který tíží většinu poskytovatelů internetových služeb.

Problém detekce anomálního chování uživatelů získává v posledních letech na důležitosti, a to ruku v ruce s tím, jak se metody strojového zpracování stávají základní součástí většiny internetových služeb. Řada internetových aplikací dynamicky upravuje svůj obsah a funkcionalitu na základě četnosti přístupů, kliků. Nekorektní uživatelé se snaží získat výhodu umělým vytvářením přístupů k vlastním zdrojům a tím službu (resp. algoritmy v pozadí) “donutit” prezentovat je jako důležité. Díky vynalézavosti takových uživatelů, anonymity z pohledu poskytovatele služby a velké variabilitě v chování uživatelů se úloha automatické detekce takového podvodného chování stává velice náročnou. Tato bakalářská práce si kladla za cíl poskytnout přehled existujících metod, vybrat vhodné řešení, implementovat jej a vyhodnotit jeho kvalitu.

V úvodních kapitolách je definován problém, jednotlivé pojmy, které se v této oblasti používají, a uvádí řadu příkladů, ze kterých vyplývá významnost daného problému z pohledu poskytovatelů internetových služeb. Provedl rozsáhlou rešerši vědecké literatury s cílem sumarizovat současná řešení, na jejichž základě postavil řešení vlastní. V následujících kapitolách postupně uvádí využití technologie, charakter a vlastnosti dat a detailní popis navrhovaného řešení. V následujících kapitolách jsou uvedeny dosažené výsledky, návrh kroků pro budoucí vývoj a závěr shrnující dosažené cíle.

Tomáš Vyskočil přistupoval k řešení úkolu po celou dobu velice pragmaticky a s vysokým nasazením. Celý návrh a realizace řešení vyžadovaly zpracování rozsáhlého množství dat s velkými nároky na množství potřebných výpočtů. Student diplomě s tímto “problémem” vypořádal využitím technologií pro zpracování velkých dat (Hadoop/MapReduce), které byly nad rámec současné výuky. Je nutné zdůraznit, že předkládané řešení není pouhou aplikací existujících metod, ale je šité na míru problému, obsahu a rozsahu poskytnutých dat. Při volbě modelů také správně zohlednil efektivnost a škálovatelnost vybraného řešení. Student se také velice dobře vypořádal s vyhodnocením svého řešení. Přestože měl k dispozici pouze neanotovaná data, tak na základě diskuze s poskytovatelem dat a vlastním návrhem generování pochybného chování, dokázal poskytnout pohled do efektivity vlastního řešení.

Bakalářská práce má všechny atributy kvalitní technické zprávy. Je třeba zdůraznit, že student předkládá originální práci, která je založena na znalosti celé šíře problematiky, a to zpracování rozsáhlých souborů protokolů, návrh modelů chování, extrakce a využití signálů, které generuje infrastruktura, atp.

Práce je systematicky a přehledně členěna, včetně seznamu literatury, obrázků a tabulek. Číslování kapitol a sekcí odpovídá běžným standardům. Obsahuje seznam použité literatury, na kterou se text pravidelně odkazuje. Obrázky a tabulky jsou citlivě včleněny do textu, vhodně ilustrují popisovanou problematiku a výsledky experimentů. Celá práce má správný a logický sled. Je třeba ocenit, že je napsána v anglickém jazyce, což je pro tento typ tematiky velmi vhodné. V textu se autor nevyhnul některým zvláštním formulacím, které ale nesnižují srozumitelnost a kvalitu předložené práce.

Na závěr zbývá jen konstatovat, že cíle bakalářské práce byly beze zbytku splněny. Bakalářská práce obsahuje analýzu, návrh a úplnou implementaci celkem 3 modelů pro detekci anomálního chování uživatelů spolu s kritickou analýzou jejich kvality z pohledu zadaných metrik. Předkládaná práce překračuje svou kvalitou obvyklé nároky na bakalářskou práci. Splňuje veškeré požadavky na ni kladené. Doporučuji předkládanou bakalářskou práci k obhajobě a hodnotím ji známkou **A - výborně**.

v Praze 8.6. 2015

Jan Šedivý