

Návrh zadania diplomovej práce

Predbežná verzia (podpísaná) ¹

Študent:

Meno, priezvisko, tituly: Matúš Cimerman, Bc.
Študijný program: Informačné systémy
Kontakt: matus.cimerman@gmail.com

Výskumník:

Meno, priezvisko, tituly: Jakub Ševcech, Ing.

Projekt:

Názov: Analýza prúdu prichádzajúcich udalostí použitím rôznych metód pre analýzu údajov
Názov v angličtine: Stream analysis of incoming events using different data analysis methods
Miesto vypracovania: Ústav informatiky a softvérového inžinierstva, FIIT STU, Bratislava
Oblasť problematiky: analýza dát, prúd udalostí

Text návrhu zadania²

Analýza údajov v sebe spája rôzne techniky z rôznych oblastí štatistiky, strojového učenia a dolovania v údajoch v spojení so znalosťou domény. Každá z týchto domén vyžaduje netriviálne znalosti potrebné pre preloženie otázky do analytickej úlohy, výberu analytickej metódy, vykonanie analýzy a interpretovanie výsledkov. Často je veľmi náročné nájsť experta, ktorý by vedel prepojiť všetky tieto oblasti a s ďalším rozvojom analytických metód bude tento problém ďalej rásť.

Priestor na zmiernenie tohto problému je napríklad v návrhu nástrojov, ktoré pomáhajú v tomto procese a umožňujú používanie a interpretovanie pokročilých modelov doménovým expertom bez potreby detailných znalostí o fungovaní modelu. Podobné prístupy sme mohli vidieť v podobe rôznych populárnych nástrojov na spracovanie statických kolekcii údajov pomocou metód ako sú lieviová analýza (angl. funnel analysis) alebo vnáranie sa (angl. drill down). V súčasnosti sa však do pozornosti dostáva analýza údajov v čase ich vzniku, kde hovoríme o analýze prúdu prichádzajúcich udalostí.

Analyzujte možnosti použitia známych metód na analýzu statických kolekcii údajov a existujúcich metód na analýzu prúdov údajov. Vyberte a aplikujte metódu na analýzu údajov v doméne spracovania prúdov údajov. Sústreďte sa pritom na použiteľnosť metódy, jej jednoduchosť a interpretovateľnosť poskytnutých výsledkov používateľom, ktorí nemajú detailné znalosti o fungovaní modelu. Navrhnuté riešenie overte pomocou softvérovej súčiastky implementovaním vybranej metódy vhodnej pre analýzu prúdu udalostí vo zvolenej doméne.

¹ Vytlačiť obojstranne na jeden list papiera

² 150-200 slov (1200-1700 znakov), ktoré opisujú výskumný problém v kontexte súčasného stavu vrátane motivácie a smerov riešenia

Literatúra³

- KREMPL, Georg, et al. Open challenges for data stream mining research. ACM SIGKDD Explorations Newsletter, 2014, 16.1: 1-10.
- GABER, Mohamed Medhat; ZASLAVSKY, Arkady; KRISHNASWAMY, Shonali. Mining data streams: a review. ACM Sigmod Record, 2005, 34.2: 18-26.

V Bratislave dňa 10.12.2015

Podpis študenta

³ 2 vedecké zdroje, každý v samostatnej rubrike a s údajmi zodpovedajúcimi bibliografickým odkazom podľa normy STN ISO 690, ktoré sa viažu k téme zadania a preukazujú výskumnú povahu problému a jeho aktuálnosť (uvedte všetky potrebné údaje na identifikáciu zdroja, pričom uprednostnite vedecké príspevky v časopisoch a medzinárodných konferenciách)