

# Posudek oponenta diplomové práce

Autor/autorka práce: **Michal Smeták**

Název práce: **Vizualizace dat a mashup aplikace (2013)**

Michal Smeták se ve své diplomové práci zabývá moderními způsoby využití dat. Primárně jsou uvažována data pocházející z České tiskové kanceláře, a to dvojího typu – dokumenty a události. Události jsou interní neustále se vyvíjející záznamy o budoucím a současném dění, používají se pro propojování dokumentů a ke generování plánů a dalších automaticky tvořených dokumentů. Dokumenty jsou textové zprávy, fotky, grafy, zvuky a videa, které různými distribučními cestami putují k odběratelům.

Autor DP se měl nejdříve seznámit s používáním metadat v ČTK a analyzovat veřejně dostupné zdroje dat. Z textu vyplývá, že rozboru metadat byla věnována dostatečná pozornost, součástí kapitoly dvě je popis i grafické znázornění poměrně komplikované struktury dokumentů i událostí. Zdrojům dat na Internetu bylo také věnováno dost pozornosti, autor se zaměřil především na geografickou vizualizaci a velmi podrobně popsal tomu odpovídající nástroje, co se týče zdrojů dat na webu, tak nechybí popis toho, co je možné nalézt u ČSÚ, zajímavý portál je také zmiňovaný server EPD, kde, jak se správně zmiňuje, uživatel často narazí na nedostupnost souborů. Kromě mapových nástrojů se autor ještě významně věnuje zobrazování dat na časových řadách, což pokládám za užitečné.

Kapitoly šest a sedm jsou potom věnovány praktické části DP. Jsou popsány způsoby, jak dostat vlastní data na mapový podklad, rozebrány použité nástroje. Velmi podrobně je řešeno shlukování geografických souřadnic, jsou popsány shlukovací algoritmy, možné způsoby implementace jak na serverové části tak na straně klienta. Autor ukazuje postup vývoje od návrhu přes vlastní vývoj až po testování. Serverová část je dodána jako webový modul (war), který stačí nahrát na aplikační server Tomcat. V části pro testování je věnován největší prostor náročnosti shlukovacího algoritmu, jsou připraveny tabulky a grafy ukazující, jak se výpočetní (a tedy i časová) náročnost zvyšuje s počtem značek a úrovní zoomu, je uvažována i varianta nejhorší, kdy k žádnému seskupení nedojde.

Dosažené výsledky mi připadají uspokojivé, z hlediska potřeb ČTK splňují očekávání. Autorem vybrané technologie jsou použitelné, nenarazí se na žádné licenční omezení. Dobrou volbu potvrzuje i použití zvolené technologie Leaflet serverem Ihned.cz v sekci datablogů. Doba odezvy v případě seskupování pro plánované počty najednou zobrazovaných záznamů by měla bohatě stačit. Zobrazování časové osy se zdá také vyhovující. V práci mi trošku chybí zmínka o jiných způsobech vizualizace dat - grafech, dvojrozměrné či trojrozměrné žebříčky (tag cloud) ukazující četnost nejčastěji používaných klíčových slov, trendů apod.

Z formálního hlediska mi práce připadá v pořádku, narazil jsem pouze na několik málo překlepů a všiml jsem si pouze jedné významnější gramatické chyby u konce. DP je dobře strukturovaná, obsahuje všechny formální prvky, obrázky i přílohy jsou patřičně označeny, použitá literatura také. Autor svou práci dokazuje, že se orientuje v odborné literatuře, je schopen nastudovat informace týkající se dané problematiky. Stojí za zmínku, že většina použité literatury pochází z webu, v seznamu zdrojů jsou uvedeny URL adresy prakticky u všech záznamů.

Jednotlivé body zadání autor splnil, pouze jsem nenašel návrhy dalších možných rozšíření (zadání, bod 4.) a jak bylo výše uvedeno, v práci by mohlo být popsáno více různých nástrojů a způsobů

**SOUHLASÍ  
S ORIGINÁLEM**



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

vizualizace, na druhou stranu velice dopodrobna až skoro nad rámec práce byla popsána problematika zobrazování geografických dat.

Dotazy k práci

- Zkuste se zamyslet nad dalším rozšířením práce, co by podle Vás stálo za podrobné zkoumání?
- Bylo by možné doplnit do mapových podkladů demografická data – např. zobrazovat počty zpráv s kategorií krimi či finance z jednotlivých krajů v závislosti na věku, pohlaví, zaměstnanosti v daném kraji, jak byste to realizoval?

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

V Praze 3.6.2013

Ing. Jan Pluskal



**SOUHLASÍ  
S ORIGINÁLEM**



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

①