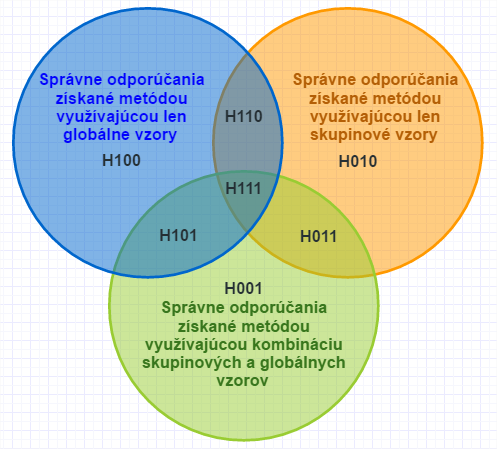
# PRÍLOHA F – Dodatočné materiály k výslednému vyhodnoteniu pre dataset ALEF

V tejto časti uvádzame niektoré podrobnejšie grafy a tabuľky z celkového vyhodnotenia pre dataset ALEF.

## Vyhodnotenie vzťahov medzi globálnymi a skupinovými vzormi [dodatočné materiály]

Uvádzame Tabuľka 1, ktorá sumarizuje zaujímavé fakty získané z rôznych prienikov množín vektorov úspešnosti odporúčania (pre odporúčania získané metódy využívajúcou len globálne vzory, len skupinové vzory a kombináciu skupinových vzorov s globálnymi) a pre najlepšiu nájdenú konfiguráciu pre dataset ALEF. Množiny sú označené písmenom *H* nasledovaným troma bitmi, ktoré označujú použitie globálnych, skupinových vzorov a ich kombinácie. Rôzne prieniky sú pre pripomenutie znázornené na Vennovom diagrame (Graf 1).



Graf 1 Vennov diagram znázorňuje označenia jednotlivých množín vektorov úspešnosti odporúčania a ich prienikov.

Na grafe 2 uvádzame distribúciu počtu používateľov v jednotlivých skupinách počas spracovania prúdu dát. Na grafe 3 uvádzame počty unikátnych vzorov v jednotlivých skupinách počas spracovania prúdu dát.

Tiež uvádzame príklad nájdených top vzorov v jednotlivých skupinách v niektorých snímkach v Tabuľka 2. Ilustrujeme tak, že dolovanie skupinových vzorov v prúde dát ponúka zaujímavý pohľad na správanie a zmeny správania používateľov v skupinách.

V ostatných grafoch na konci prílohy uvádzame priebehy vývoja pre metriku *presnosť* pre 2 odporúčané položkyv jednotlivých skupinách počas spracovania prúdu dát. Nech GGC je označenie pre metódu využívajúcu globálne aj skupinové vzory, GO označenie pre metódu využívajúcu iba globálne vzory a OG metódu využívajúcu iba skupinové vzory.

Pre všetky uvedené grafy a tabuľky platí, že merania sú uskutočňované v časových snímkach. Časové snímky predstavujú určitý bod v čase počas spracovania prúdu dát, ktorý sa pravidelne opakuje. Konkrétne je to tesne pred ďalším makrozhlukovaním (vtedy by už malo byť dosť používateľov zaradených do skupín a mali by byť identifikované dostatočne silné aktuálne vzory).

Graf Distribúcia počtu používateľov v skupinách v jednotlivých časových snímkach. Na osi x sú uvedené časové snímky.Snímka predstavuje moment v behu programu tesne pred makrozhlukovaním keď sa zozbierali uvedené dátata.

Graf Počty unikátnych vzorov v skupinách zaznamenané v jednotlivých časových snímkach na logaritmickej škále (kvôli lepšej prehľadnosti). Na osi x sú uvedené časové snímky.Snímka predstavuje moment v behu programu tesne pred makrozhlukovaním keď sa zozbierali uvedené dáta.

Tabuľka 1 Suma rôznych prienikov vektorov správnosti odporúčania získaných pomocou globálnych skupinových aj kombinovaných vzorov.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***ID*** | ***POPIS*** | ***Spôsob výpočtu*** | **P@1** | **P@2** | **P@3** | **P@4** | **P@5** | **P@10** | **@15** |
| **1** | Počet správnych odporúčaní len pomocou skupinových vzorov a nie globálnych: | **H011 + H010** | **5.86%** | **7.17%** | **7.97%** | **8.33%** | **8.82%** | **9.96%** | **10.27%** |
| **2** | Počet správnych odporúčaní len pomocou globálnych vzorov a nie skupinových: | **H101+H100** | **22.72%** | **24.09%** | **25.19%** | **25.82%** | **26.31%** | **26.98%** | **27.32%** |
| **3** | Počet správnych odporúčaní len pomocou kombinovaných vzorov a nie iných vzorov: | **H001** | **0.32%** | **0.84%** | **1.19%** | **1.54%** | **1.84%** | **2.85%** | **3.49%** |
| **4** | Počet správnych odporúčaní len pomocou skupinových vzorov, ktoré kombinovaná metóda nevedela odporučiť | **H010** | **2.63%** | **3.41%** | **3.95%** | **4.16%** | **4.40%** | **5.02%** | **5.08%** |
| **5** | Počet správnych odporúčaní pomocou globálnych vzorov, ktoré kombinovaná metóda nevedela odporučiť | **H100** | **2.02%** | **2.60%** | **3.00%** | **3.27%** | **3.47%** | **4.05%** | **4.34%** |
| **6** | Počet správnych odporúčaní získaných pomocou prieniku odporúčaní z globálnych a skupinových vzorov. | **H111+H110** | **26.55%** | **25.77%** | **25.24%** | **24.91%** | **24.74%** | **24.12%** | **23.13%** |
| **7** | Teoreticky najvyšší možný počet správnych odporúčaní, ktoré by mohla kombinovaná metóda dosiahnuť. | **H111+H110+H101+H100+H011+H010** | **55.14%** | **57.02%** | **58.40%** | **59.06%** | **59.88%** | **63.91%** | **60.72%** |
| **8** | Počet správnych odporúčaní pomocou našej metódy | **H101+H011+H001+H111** | **50.71%** | **51.26%** | **51.73%** | **51.96%** | **52.41%** | **52.62%** | **52.13%** |
| **9** | Počet správnych odporúčaní pomocou globálnych vzorov | **H111+H110+H101+H100** | **49.27%** | **49.85%** | **50.44%** | **50.73%** | **51.06%** | **51.10%** | **50.45%** |
| **10** | Počet správnych odporúčaní pomocou skupinových vzorov | **H111+H011+H110+H010** | **32.41%** | **32.93%** | **33.21%** | **33.24%** | **33.57%** | **34.09%** | **33.40%** |

Tabuľka Ukážky vybraných najsilnejších globálnych a skupinových vzorov vo vybraných snímkach. -1 označuje globálne vzory.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | ID SNÍMKY | | |
| SKUPINA |  | **7 (2014-03-01 14:25:11)** | **8(2014-04-03 14:09:06)** | **9 (2014-06-17 15:10:58)** |
| **-1** | [1.3 Softvér],[1.1 Čo je to softvérové inžinierstvo?] | [5.1.4 Diagram modelu údajov], [5.1.3 Entitno-relačný diagram] | [FLP:Predikát op], [Operátory] |
| [1.4 Dôležité skutočnosti k vývoju softvéru],[1.1 Čo je to softvérové inžinierstvo?] | [5 Štruktúrne modely], [4 Funkcionálne modely] | [3.6 UML diagramy], [3 Modelovanie softvéru] |
| **0** | [5 Štruktúrne modely],[6 Modely správania] | [4 Funkcionálne modely],[4.2.2 Notácia pre model prípadov použitia] | [4.2 Model prípadov použitia],[4.2.1 Použitie modelu prípadov použitia] |
| [PSI - question 4], [PSI - question 144] | [4 Funkcionálne modely],[4.2.1 Použitie modelu prípadov použitia] | [5 Štruktúrne modely],[4 Funkcionalne modely] |
| **1** | [5.1 Model údajov],[5.1.2 Základné elementy modelu údajov],[5.1.3 Entitno-relačný diagram] | [1.4.2 Dôsledky a príčiny],[1.4.1 Zákonité a nezákonité problémy] | [1.4.2 Dôsledky a príčiny],[1.4.1 Zákonité a nezákonité problémy] |
| [5.1 Model údajov],[5.1.1 Typy modelov údajov] | [1.4.1 Zákonité a nezákonité problémy] | [1.4.1 Zákonité a nezákonité problémy] |
| **2** | [1.2 Softvérové inžinierstvo - dôležité pojmy] | [4.2.1 Použitie modelu prípadov použitia],[4.2.2 Notácia pre model prípadov použitia] | [3.1 Model softvérového systému] |
| [FLP:Príklad SUCIN], [FLP:Základné prvky jazyka lisp] | [5 Štruktúrne modely], [4 Funkcionálne modely] | [3.3 Rôzne pohľady - rôzne modely softvéru] |
| **3** | [5 Štruktúrne modely],[6 Modely správania] | [FLP:Funkcionál REDUCE],[FLP:Funkcionál MAPCAR],[FLP:Funkcionál FUNCALL] | [FLP:Funkcionály FIND-IF a COUNT-IF],[FLP:Funkcionál MAPCAR],[FLP:Funkcionál FUNCALL] |
| [4 Funkcionálne modely],[6 Modely správania] | [FLP:Funkcionál REDUCE],[FLP:Funkcionál MAPCAR],[FLP:Forma FUNCTION] | [FLP:Funkcionál REDUCE],[FLP:Funkcionál MAPCAR],[FLP:Funkcionál FUNCALL] |
| **4** | [1.3 Softvér],[1.1 Čo je to softvérové inžinierstvo?] | [FLP:Funkcionál REDUCE],[FLP:Funkcionál MAPCAR],[FLP:Forma FUNCTION],[Funkcionál FUNCALL],[Funkcionály REMOVE-IF a REMOVE-IF-NOT] | [FLP:Počítanie],[FLP:Redukcia],[FLP:Zobrazenie] |
| [1.4 Dôležité skutočnosti k vývoju softvéru],[1.1 Čo je to softvérové inžinierstvo?] | [FLP:Funkcionál REDUCE],[FLP:Funkcionál MAPCAR],[FLP:Forma FUNCTION],[Programové schémy],[Funkcionál FUNCALL],[Funkcionály REMOVE-IF a REMOVE-IF-NOT] | [FLP:Počítanie],[FLP:Filter],[FLP:Zobrazenie] |
| **5** | [3.1 Model softvérového systému],[3.2 Aspekty modelovania softvéru] | [3.1 Model softvérového systému],[3.2 Aspekty modelovania softvéru] | [5 Štruktúrne modely], [6 Modely správania] |
| [5 Štruktúrne modely],[6.3.1 Použitie stavového diagramu],[PSI - question 7], [PSI - question 2], [PSI - question 44] | [5 Štruktúrne modely],[6.3.1 Použitie stavového diagramu],[PSI - question 7], [PSI - question 2], [PSI - question 44] | [5 Štruktúrne modely] |