

Odporúčanie posúdenia zdrojového kódu

Autor: Bc. Matej Chlebana
(matej.chlebana@gmail.com)

Vedúci: Ing. Karol Rástočný



Cieľ práce

- Zvýšenie efektivity prehliadok zdrojového kódu
- Zvýšenie kvality softvéru
- Využitie dolovania procesov pri identifikácii chybových zdrojových kódov
- Overenie možnosti využitia dolovania procesov pri identifikácii autorov zdrojového kódu

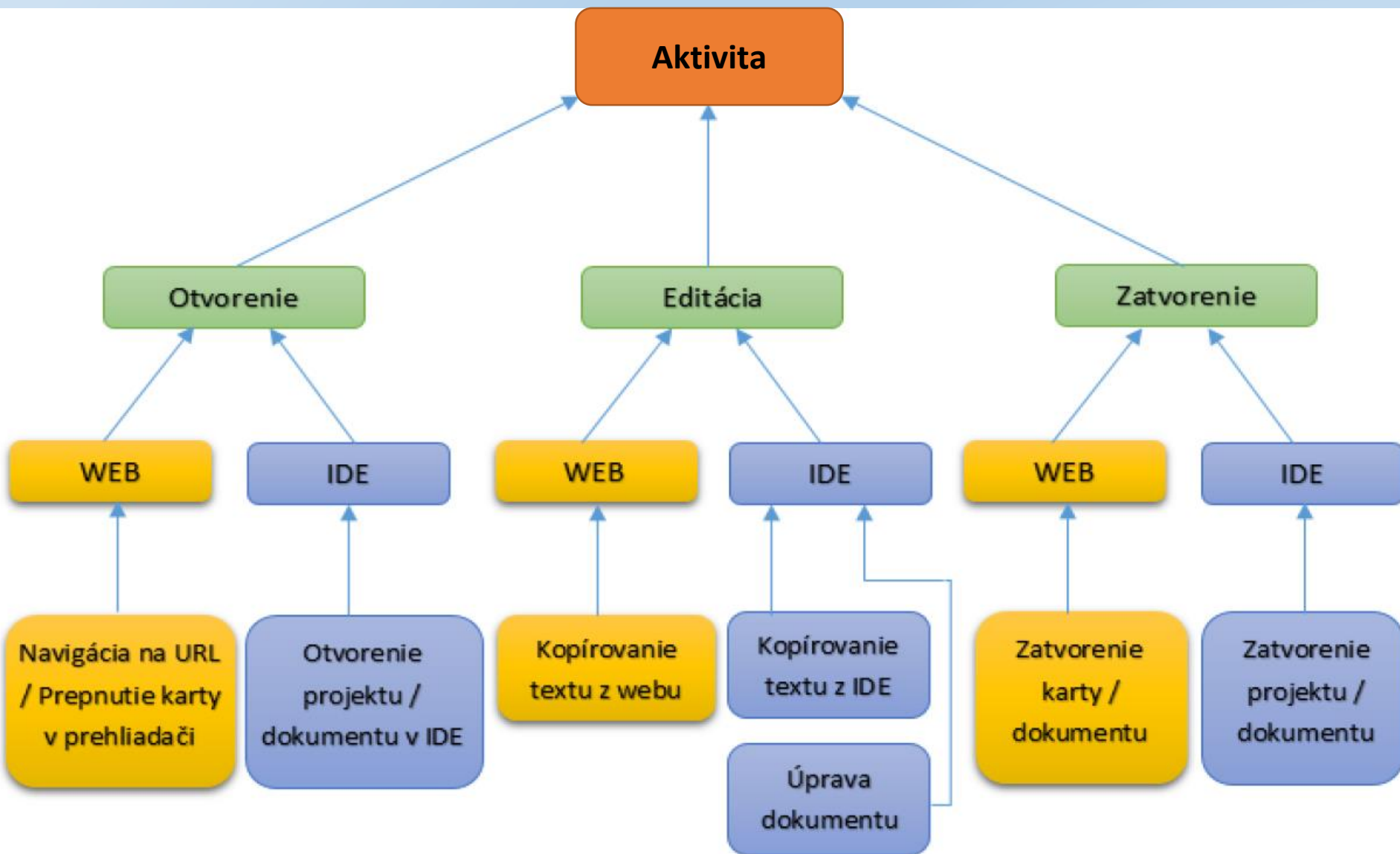
Prehliadky zdrojového kódu

- Statické testovanie – manuálna kontrola kódu
- Efektívnosť prehliadky: 55 – 60 % [McConnel, 2004]
nájdenných chýb
 - Jednotkové testovanie – 25%
 - Funkčné testovanie – 35%
 - Integračné testovanie – 45 %
- Zvyšovanie kvality produktu
- Náklady: 5 – 15 % prostriedkov projektu

Sledovanie práce vývojárov

- Nový pohľad na vznik novej verzie zdrojového kódu
 - Postupnosť aktivít vývojára
- Zaznamenávanie aktivít – projekt PerConIK
 - IDE – písanie kódu, kopírovanie / vkladanie kódu, práca s dokumentmi...
 - Web – navigácia na URL, práca so záložkami...
 - Rozdelenie podľa navštívenej domény – štúdium, zábava, vyhľadávanie, testovanie, súkromné...
 - BASH príkazy
- Sledovanie práce vývojárov vo firme Gratex
- Taxonómia aktivít – rôzne úrovne granularity

Taxonómia aktivít

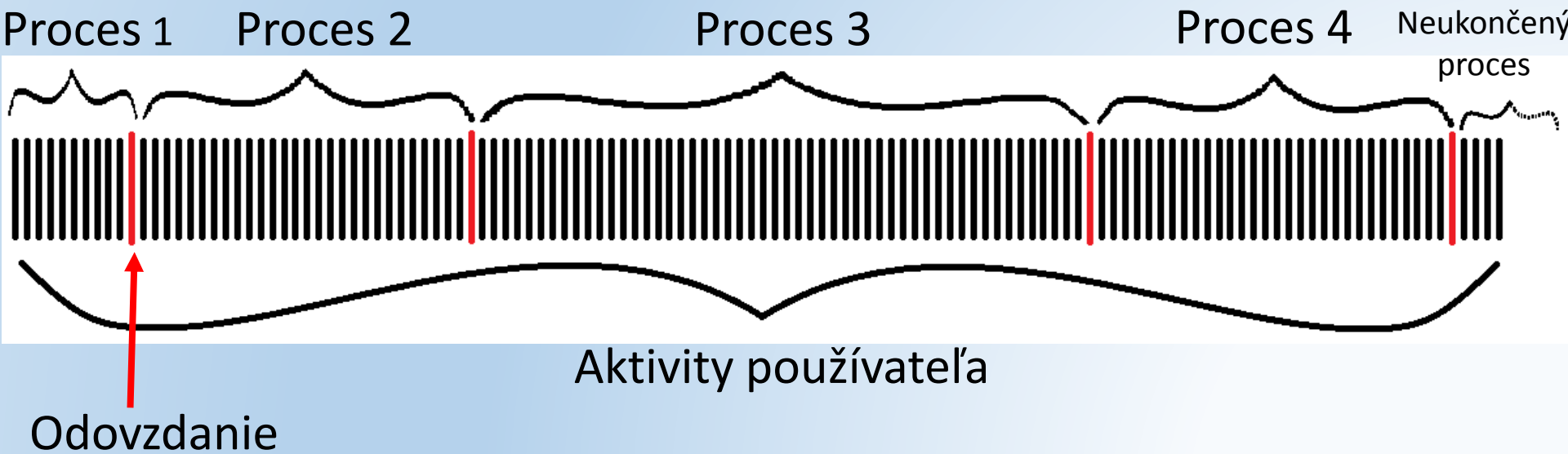


Rozdelenie - navigácia na URL

- Štúdium - *blog, community, microsoft, java, wiki, support...*
- Zábava - *topky, nay, twitter, facebook, youtube, video...*
- Vyhľadávanie - *google, bing, search...*
- Testovanie - *localhost, :8080, :8081*
- Súkromie - *ib.slsp, ib.vub, mail., gmail*
- Lokálne - *gratex, dropbox, github*

Procesy

- Sekvencia aktivít rozdelená odovzdaniami (commit)
- Jeden proces – jeden používateľ
- Rôzna zložitosť procesov



Anotácia procesov

- Anotácia zdrojového kódu vytvoreného v procese
 - Informačné značky – Bug, Smell, Fix, Ranking...
 - Extrakcia anotácií z dokumentačných komentárov pri odovzdaní novej verzie
- Identifikácia povahy procesu
 - Chybový, opravný, neutrálny...

CodeReview



Anotácia procesov

Vytvorenie procesov

Generovanie .XES súboru



Process Miner

ITM



CORD



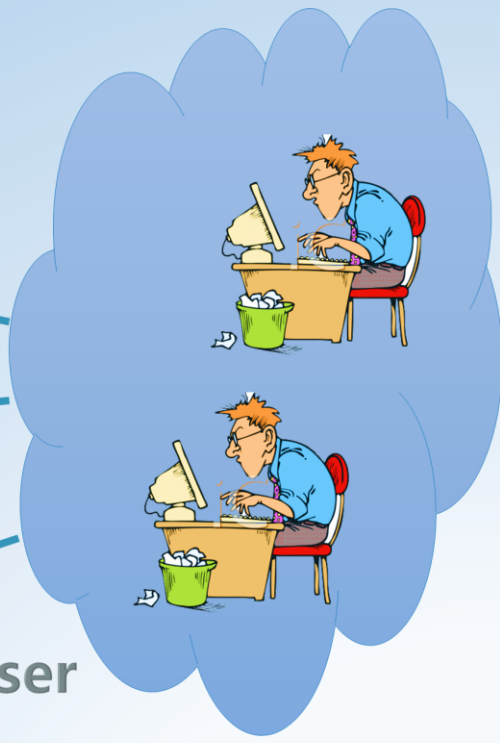
UACA

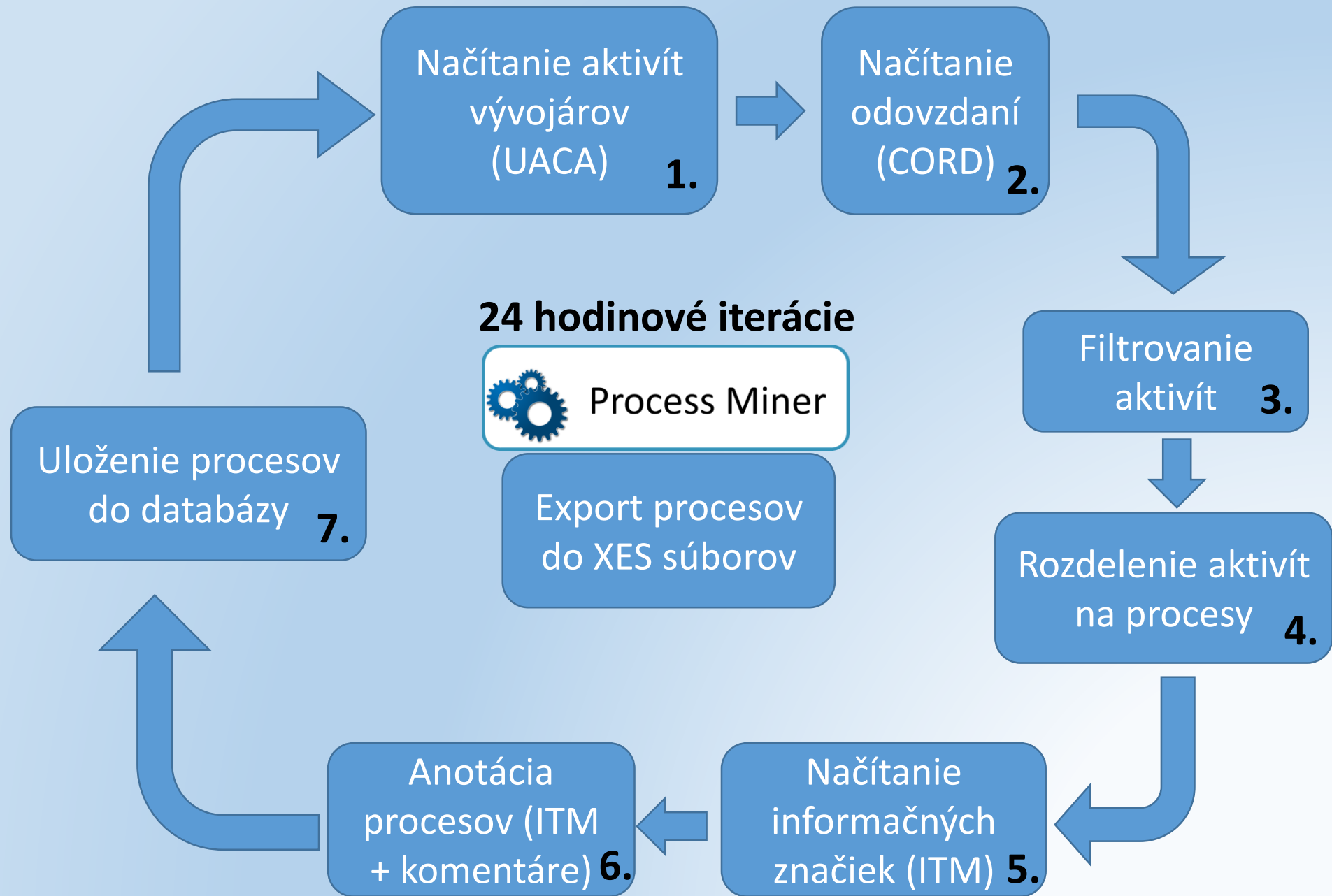


Bug Fix
Smell Ranking
Informačné značky

Odovzдания

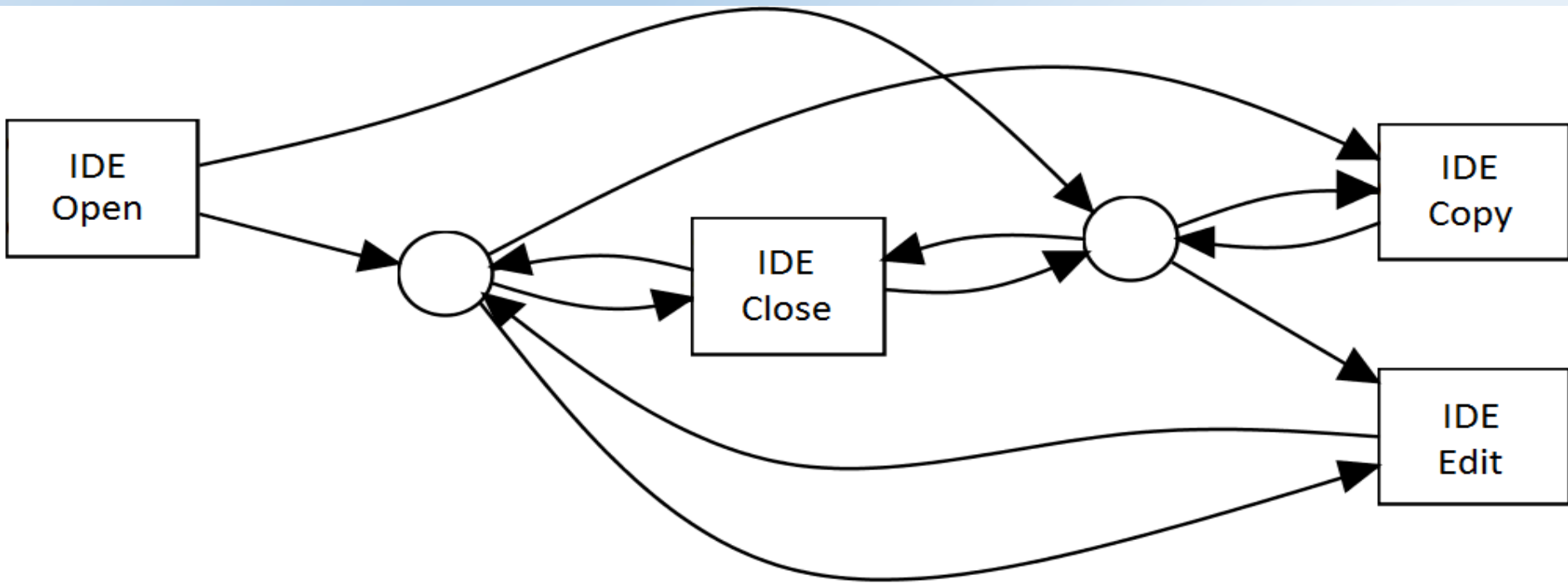
Aktivity
IDE Web Browser
Copy Bash
Paste

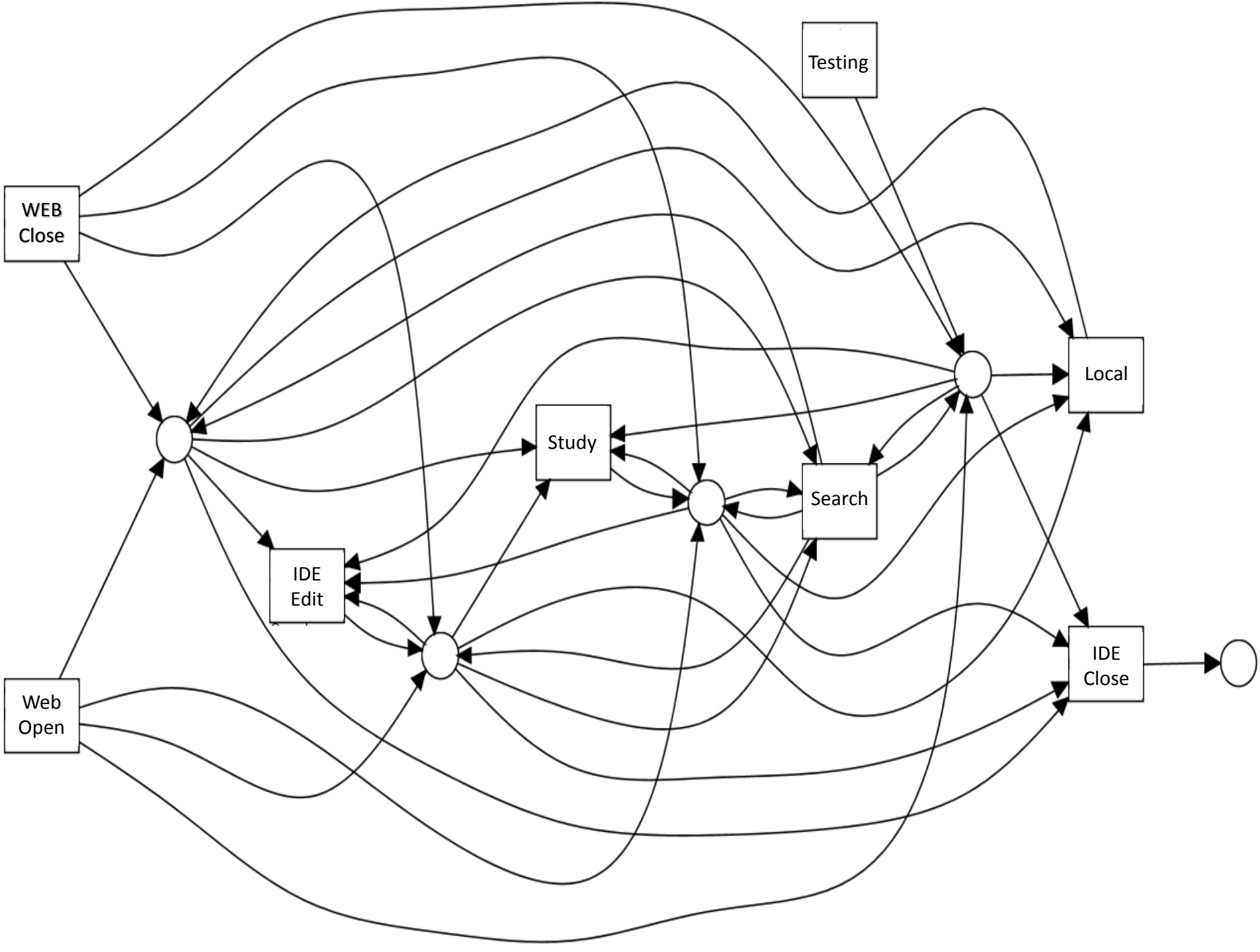




Dolovanie procesných modelov

- Vytvorenie procesných modelov v programe RapidMiner s rozšírením ProM
- Vstup – XES súbor





Dolovanie procesných modelov

- Natrénovanie procesného modelu – Petriho sieť
 - Identifikácia chybových procesov
 - Pre jednotlivých vývojárov
 - Nezávisle od vývojára
 - Identifikácia autora procesu
 - Rôzna povaha procesov pre jednotlivých vývojárov
- Porovnanie procesného modelu so záznamom aktivít
 - Finess: Trace, Move-Model, Move-Log
 - Vyhodnotenie klasifikácie: presnosť, pokrytie a F1

Identifikácia chybových procesov

- Odporúčanie posúdenia chybových zdrojových kódov vytvorených počas chybového procesu
- Merania pre hranice podobnosti 0,1 – 0,95
- Efektívnosť prehliadky 55 – 60 %
- Smerodajná úroveň pokrytia pri hodnote 55 – 60 %

Prehľad početnosti dát

	Gretex International	FIIT STU
Počet používateľov UserActivity	29 (aktívnych 13)	9 (aktívnych 7)
Počet všetkých aktivít	779975 (58 dní)	183710 (84 dní)
Počet všetkých procesov	1230	277
Priemerný počet procesov na jedného aktívneho používateľa	94,6	39,6
Priemerný počet aktivít na jedného aktívneho používateľa za jeden deň	1034	312

Prehľad povahy procesov

Počet procesov	Gratex International	FIIT STU
Všetky	1230	277
Chybové	116 (10 %)	44 (16 %)
Opravné	320 (26 %)	11 (4 %)
Bez príznaku	794 (64 %)	222 (80 %)

XES súbory chybových procesov

XES súbory procesov na
identifikovanie

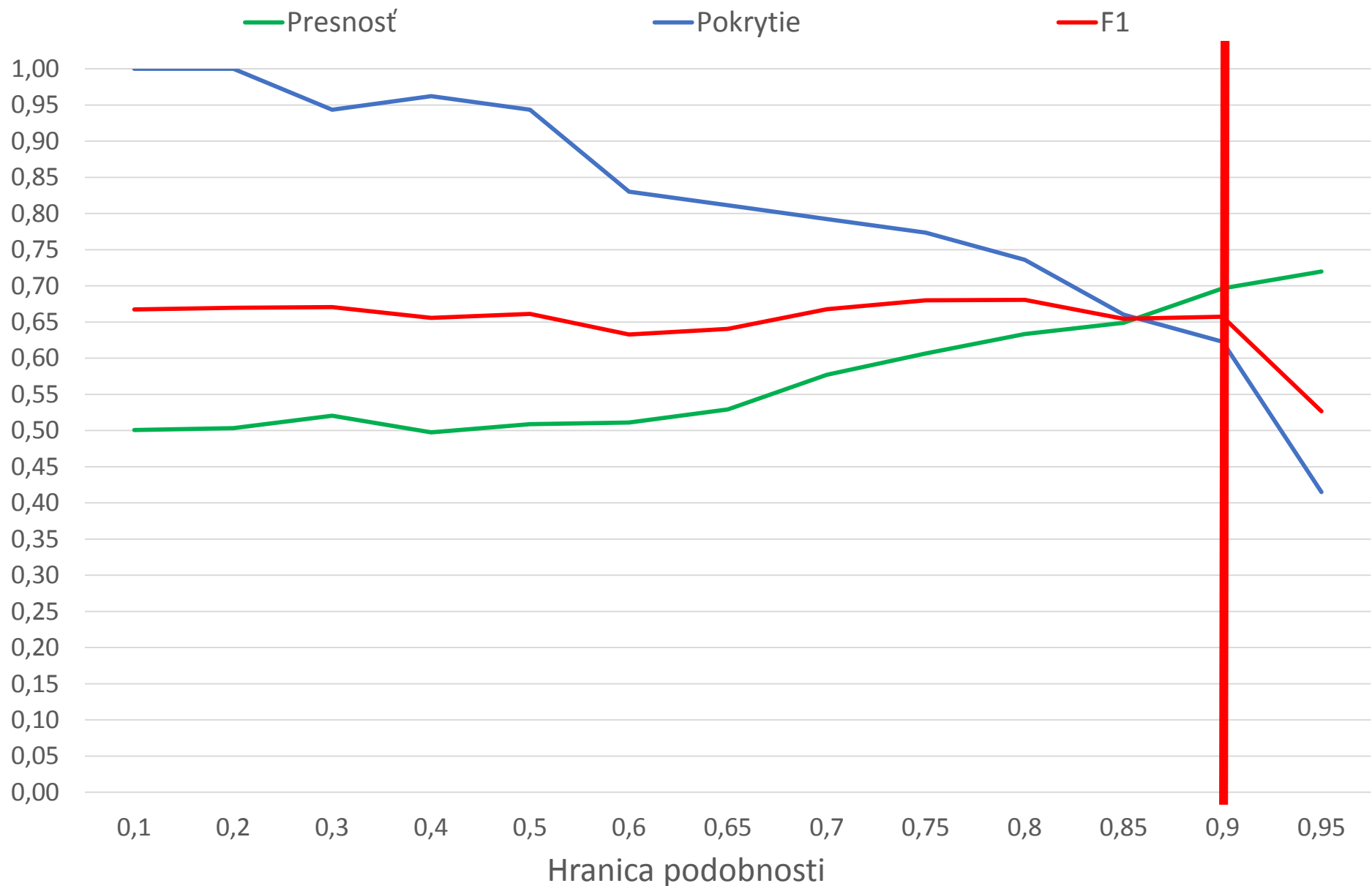
Modelovanie chybového
procesného modelu

Porovnanie

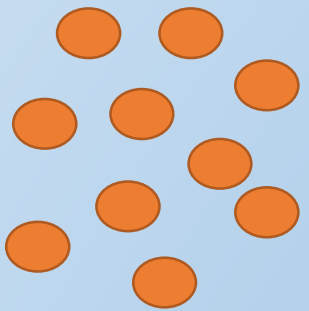
Row No.	Trace Index	Trace Fitness	Move-Log Fitness	Move-Model Fitness
1	[8, 29]	0.3125	0.33333333333333337	0.09999999999999998
2	[19, 62]	0.375	0.6666666666666667	0.18181818181818177
3	[40]	0.94895833333333333	0.9493136219640972	0.99888888888888889
4	[56]	0.7422552664188352	0.7405541561712846	0.9966101694915255
5	[39]	0.594488188976378	0.5838383838383838	1.0
6	[38]	0.4152542372881356	0.39589442815249265	0.9926470588235294
7	[14]	0.04034582132564846	0.020958083832335328	0.5384615384615384
8	[59]	0.21264367816091956	0.19104477611940296	0.9552238805970149
9	[57]	0.44690265486725667	0.4225352112676056	0.9782608695652174
10	[55]	0.19811320754716977	0.15577889447236182	0.9393939393939394
11	[51]	0.17934782608695654	0.13450292397660824	0.8846153846153846
12	[35]	0.9074074074074074	0.9194630872483222	0.9785714285714285
13	[33]	0.5612903225806452	0.5352112676056338	0.9743589743589743
14	[53]	0.971830985915493	0.9689922480620154	1.0

Identifikácia chybových procesov

Pokrytie 62 % - Presnosť 70 %



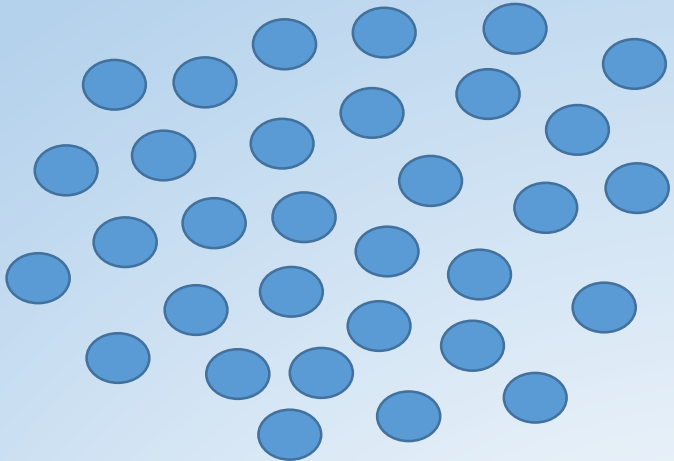
Potenciálne chybové procesy



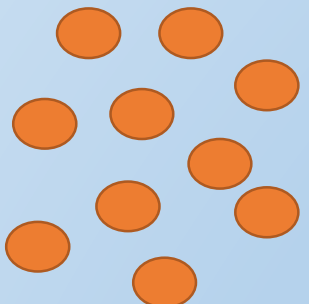
Prehliadka zdrojových kódov vytvorených počas chybových procesov

Vytvorenie nových anotácií

Správne procesy

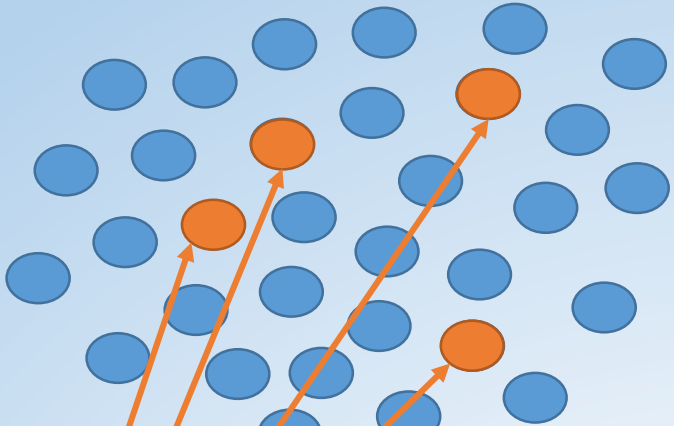


Potenciálne chybové procesy



Prehliadka zdrojových kódov vytvorených počas chybových procesov

Správne procesy

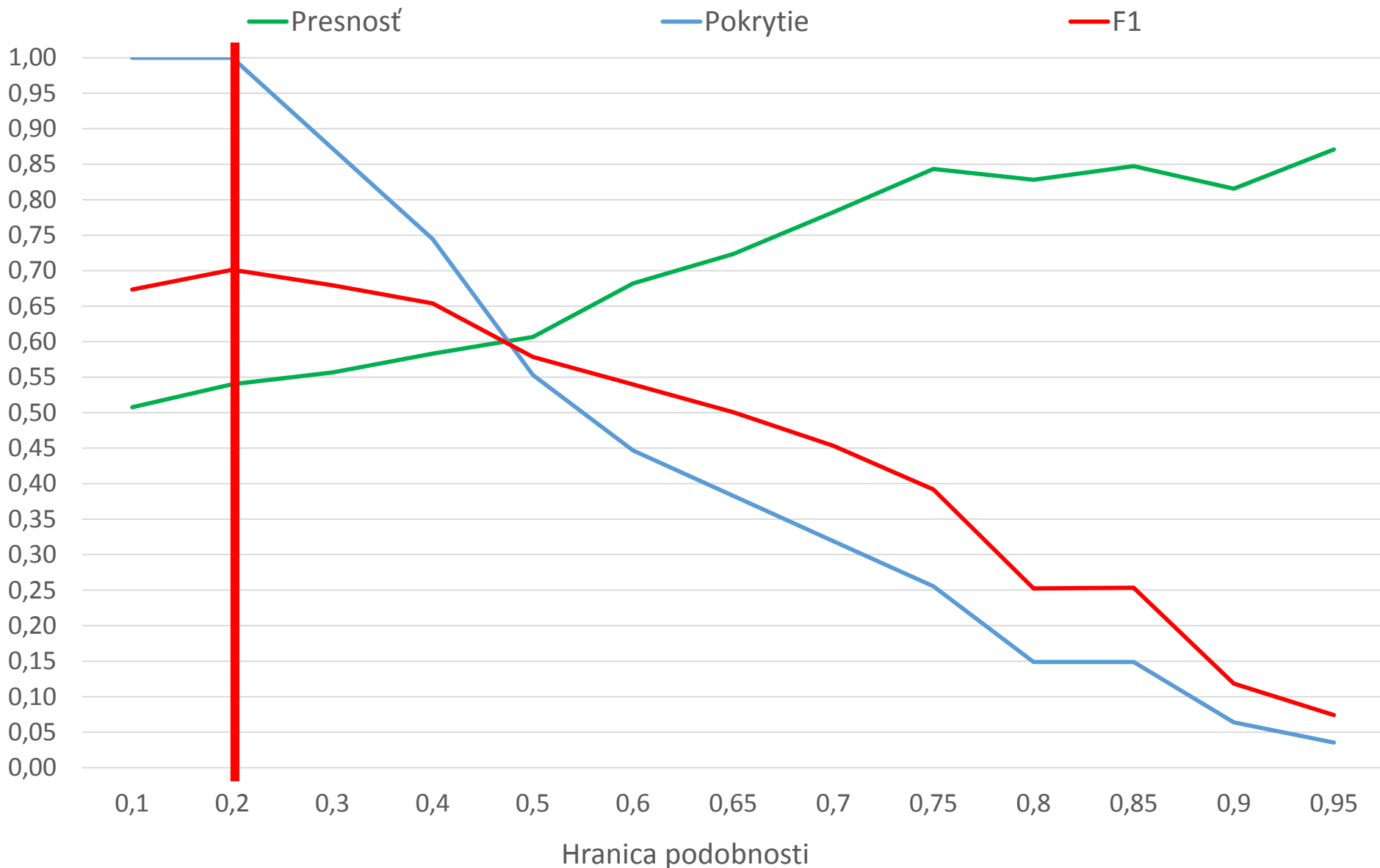


Vytvorenie nových anotácií

Identifikácia autora procesu

- Identifikácia autora zdrojového kódu
- Závislosť procesného modelu od jeho vlastníka
- Merania pre hranice podobnosti 0,1 – 0,95
- Maximalizácia metriky F1
- Testovacia množina 13 autorov
- Vlastnosť objavená pri experimentoch – nad rámec zadania

Pokrytie 100 % - Presnosť 54 % (F1 70 %)



Zhodnotenie

- Vyššia detailnosť aktivít → lepšie výsledky
 - 62% / 70% úspešnosť identifikácie chybových procesov
 - 100% / 54% úspešnosť identifikácie autora procesu
 - 95% / 56% úspešnosť identifikácie falošného autorstva
-
- Vytváranie odporúčaní pre posudzovateľov na posúdenie zdrojových kódov

Úspechy

- Publikovanie článku v zborníku IIT.SRC 2015
- Cena dekana na študentskej vedeckej konferencii IIT.SRC 2015
- Čestné uznanie na Česko-Slovenskej ŠVOČ v informatike v kategórií *Aplikovaná informatika a softvérové inžinierstvo*

Ďalšia práca

- Zvýšenie detailnosti anotácií
- Väčší počet kategórií pre klasifikáciu procesov
- Prístup ku zdrojovým kódom
- Odporúčanie posúdenia konkrétnych častí kódu

Zhrnutie

- 62% / 70% úspešnosť identifikácie chybových procesov
- 100% / 54% úspešnosť identifikácie autora procesu
- 95% / 56% úspešnosť identifikácie falošného autorstva
- Vytváranie odporúčaní pre posudzovateľov na posúdenie zdrojových kódov
- Cena dekana IIT.SRC 2015
- Čestné uznanie na Česko-Slovenskej ŠVOČ