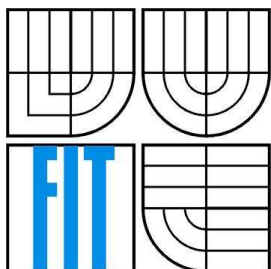




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INFORMAČNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INFORMATION SYSTEMS

METRIKY PROCESŮ VÝVOJE SOFTWARE

SOFTWARE DEVELOPMENT PROCESSES METRICS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. Jan Verner

VEDOUCÍ PRÁCE
SUPERVISOR

RNDr. Jitka Kreslíková, CSc.

BRNO 2009

Abstrakt

Pojem proces pronikl v posledních letech do všech oblastí lidské činnosti, kde pomáhá systematizovat dříve neorganizovanou práci. Tato práce se zaměřuje na analýzu procesů pomocí statistických metod (Statistical process control). V úvodu přináší základní teoretické poznatky z oblasti managementu procesů. Hlavní část je pak věnována samotným technikám, které se uplatňují při analýze dat. Jsou zde shrnuty postupy, jak jednoduše a efektivně posuzovat chování procesu pomocí grafů a statistických metod a jak následně z výsledků analýz vyvodit příslušné kroky vedoucí ke zlepšení procesů. Poslední část této práce je věnována praktické ukázce analýzy reálných dat pomocí vytvořené aplikace a rozboru dosažených výsledků, které prokazují obecnou vhodnost a použitelnost statistických metod analýzy dat.

Abstract

The importance of processes is growing. Processes are integrated into many areas of human work, where they help to organize all activities. This thesis aims on statistical approach to process behavior analysis. There are mentioned the basic process management issues in the beginning of this thesis. The main focus is on the techniques used during process analysis. There are mentioned approaches that are used to easily and effectively assess process behavior using process behavior charts and how to deduce proper steps from the analysis results to improve the process. The last part of this thesis is dedicated to practical analysis of real process behavior data using implemented solution. The results show common usability of statistical methods and approaches to process analysis.

Klíčová slova

Statistická analýza, management procesů, grafy chování procesů, proces, vspělost procesu

Keywords

Statistical analysis, process management, behavior charts, process, process capability

Citace

VERNER Jan: Metriky procesů vývoje softwaru, diplomová práce, Brno, FIT VUT v Brně, 2009

Metriky procesů vývoje softwaru

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením RNDr. Jitky Kreslíkové, CSc.

Další informace mi poskytl Mgr. Ondřej Skřehota z firmy ANF Data spol. s r. o., a Siemens Company.

Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Jan Verner
18. května 2009

Poděkování

Na tomto místě bych velmi rád poděkoval vedoucí této práce doktorce Jitce Kreslíkové za její přístup, poskytnuté znalosti a zkušenosti a za velkou ochotu při realizaci tohoto projektu. Dále Mgr. Ondřeji Skřehotovi, který ochotně poskytl data potřebná pro provedení analýz a zároveň také konzultace k dosaženým výsledkům.

Děkuji!

© Jan Verner, 2009

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

Obsah	1
1 Úvod.....	2
2 Management procesů	3
2.1 Zavádění metrik do procesů	8
2.2 Stanovení obchodních cílů	8
2.3 Definice metrik procesů	9
2.4 Získávání dat o procesech	15
3 Analýza chování procesů	19
3.1 Variabilita procesu	19
3.2 Grafy chování procesů	23
3.3 Grafy dat proměnných.....	24
3.4 Grafy kategorických dat	29
3.5 Technická stránka analýzy dat	31
3.6 Pokročilé techniky analýzy	33
4 Návrh řešení	41
4.1 Koncepce aplikace.....	41
4.2 Diagramy případů použití.....	43
5 Popis implementace	45
5.1 Základní rozdělení aplikace.....	45
5.2 Logická vrstva aplikace.....	48
6 Ukázka analýzy dat	51
7 Závěr	56

1 Úvod

Každá organizace si klade vlastní cíle. Je však otázkou, do jaké míry se jí je cíle daří plnit. Pokud jde vše podle plánů, je to v nejlepším pořádku. V opačném případě se firma dostává do problémů. Vývoj nestíhá plnit stanovené termíny, což s sebou nese zvýšení nároků na lidské zdroje, nízkou kvalitu produktů a jejich vyšší cenu. Výsledky však bývají daleko horší v okamžiku, kdy firmu začnou opouštět schopní zaměstnanci. V takovémto případě se firma dostává do spirály, ze které není úniku a která končí krachem.

Za posledních deset let výrazně stoupl význam procesů v takřka všech oblastech lidské činnosti. Původně zcela neorganizovaná práce dostává díky nim řád, stává se systematictější a tím stoupá její efektivita. Z pohledu vývoje software existuje již dlouhou dobu celá řada metodik vývoje. Mezi nejdůležitější se řadí modely vodopádu, iterativní vývoj, RUP (Rational Unified Process [IBM09] [UP]), spirálový vývoj a nejmladší metodika agilního vývoje [VAC09]. Všechny tyto zmíněné postupy jsou velmi obecné a ve všech případech nemusejí zcela vyhovovat dané organizaci.

Z tohoto důvodu je nutné některé postupy upravovat, což s sebou nese jistá rizika. Proto je nutné procesy měřit, díky čemuž můžeme sledovat jejich výkonnost, efektivitu, stabilitu a přesněji předpovídat jejich chování a výsledky. Metriky jsou důležité také tam, kde na procesy klademe vyšší nároky nebo je chceme zlepšovat. Pojmy jako management procesů a zlepšování procesů se stále častěji dostávají do slovníku manažerů a inženýrů všech společností, nehledě na jejich zaměření.

Tato práce rozebírá význam a použití metrik v procesu vývoje software se zaměřením na jeho zkvalitňování podle metodiky statistického přístupu k analýze chování procesů (Statistical process control [FLO99]). Zároveň tato práce navazuje na semestrální projekt, jehož cílem bylo seznámit se s danou problematikou a položit nezbytné teoretické základy pro další vývoj. Pozornost je věnována zejména analýze naměřených dat.

Druhá kapitola vymezuje definice několika základních pojmů, které se v oblasti managementu procesů objevují. Zdůrazňuje význam a podstatu stanovení obchodních cílů a jejich dopad na celý následující proces. Popisuje také, jakým způsobem jsou vybírány a definovány metriky procesů a uvádí nejčastější příčiny neúspěchů a možné zdroje chyb. Kapitola se také věnuje sběru dat z procesů, kdy upozorňuje na význam lidského faktoru, který je pro tuto fázi klíčový. Druhá kapitola byla řešena v rámci semestrálního projektu, na který tato práce navazuje [VER09].

Třetí kapitola přináší teoretické poznatky o analýze chování procesů. Zdůrazňuje význam stability procesu ve vztahu k jeho předvídatelnosti, popisuje příčiny variability v procesech a obsahuje nezbytné základy pro práci s grafy, které jsou hlavní součástí statistického přístupu k analýze chování procesu. Druhá část této kapitoly je věnována podrobnému popisu všech typů grafů, které se uplatňují při analýze dat. Dále zde najdeme podrobný přehled metod a technik, které se při analýze uplatňují. Závěr patří stručnému přehledu technik, které se používají při zvyšování vyspělosti procesu (*process capability*).

Čtvrtá kapitola popisuje návrh aplikace, která je součástí diplomové práce. Rozebírá základní teoretická východiska dané problematiky, nejdůležitější myšlenky aplikace a zároveň shrnuje její možnosti.

Pátá kapitola podrobně popisuje implementační stránku řešení. Zdůrazňuje význam rozdělení aplikace do dvou částí a podrobněji popisuje jednotlivé prvky systému. V závěru je jedna podkapitola věnována podrobnějšímu popisu konkrétní části aplikace.

Šestá kapitola přináší ukázkou vzorové analýzy reálných dat. Jsou zde podrobně rozebrány jednotlivé kroky analýzy a v závěru jsou shrnuty a diskutovány dosažené výsledky.

Závěr obsahuje shrnutí celé problematiky a naznačuje směr dalšího vývoje této práce.

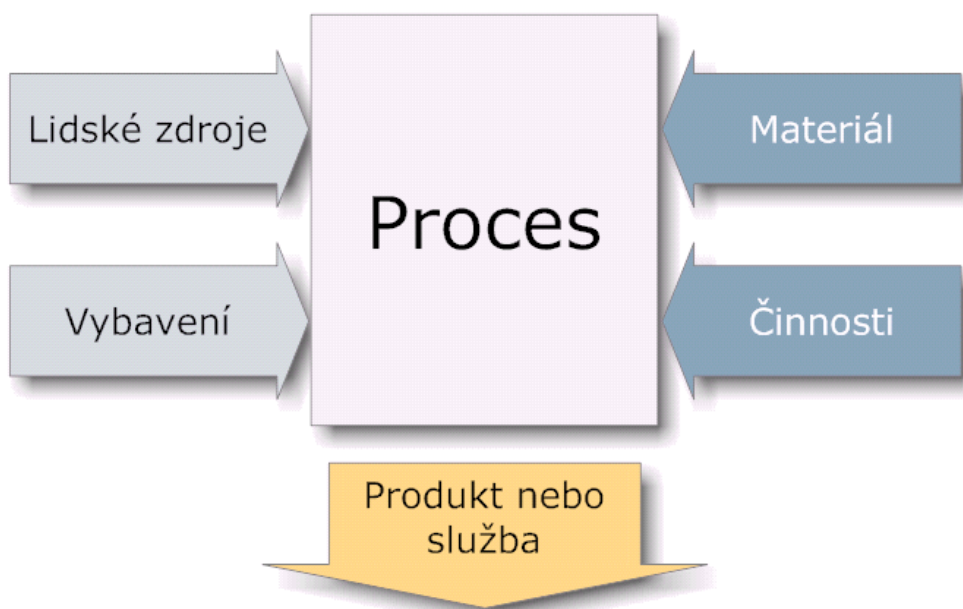
2 Management procesů

Cílem této kapitoly je zavést pojmy a obecné principy související s managementem procesů, zejména pak s jeho měřením, kontrolováním a zlepšováním. Dále tato kapitola rozebírá základní schéma měření a zkvalitňování procesů, konkrétně jeho první tři fáze. Jsou zde zmíněny postupy jak správně vytvářet definice metrik, jak shromažďovat data a uchovávat je. Kapitola také obsahuje praktická doporučení a uvádí časté příčiny chyb.

Než se budeme zabývat konkrétními kroky, které vedou ke zlepšování procesu vývoje software, je nutné nejdříve definovat základní pojmy, které budou dále používány. Obecně lze proces vývoje software, nebo-li softwarový proces, definovat mnoha způsoby [FLO99]. Pro účely této práce budeme dále softwarový proces chápat jako:

„Logické uspořádání lidských zdrojů, materiálů, vybavení a činností do pracovních aktivit, jejichž cílem je specifický produkt či služba.“

Celou situaci zachycuje Obrázek 2-1 Definice procesu.



Obrázek 2-1 Definice procesu

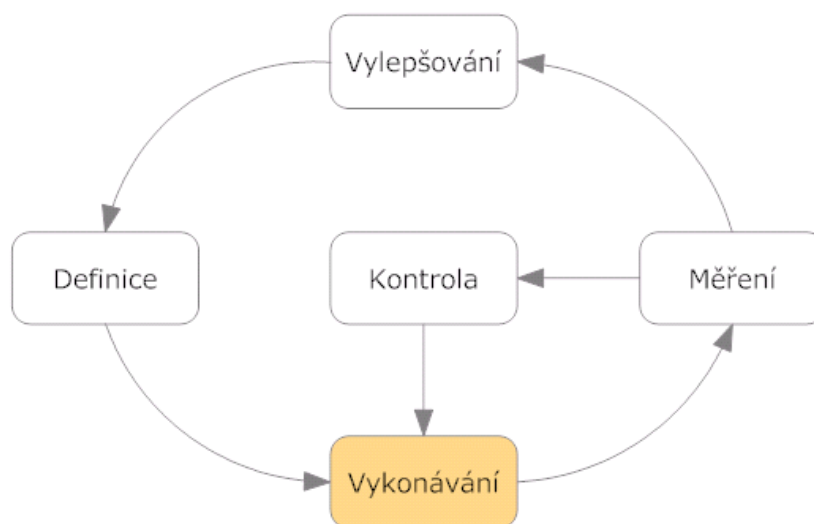
Definice procesu zohledňuje všechny faktory, které mohou mít potenciálně vliv na jeho vyspělost (capability). Při jeho zkvalitňování pak také musíme hledat slabá místa mezi výše jmenovanými součástmi. Byla by tedy chyba proces definovat jiným způsobem, jako se to například dělá v teorii operačních systémů.

Dalším pojmem je management softwarového procesu. Jeho cílem je zvládnout úkoly související s vývojem software a poskytnout zákazníkovi produkty a služby v té nejvyšší možné kvalitě a uspokojit zároveň vnitřní zájmy a potřeby firmy. Celý proces vývoje produktu nebo poskytování nějaké služby musí management chápat jako seskupení několika podprocesů. Musí zajišťovat odpovídající podporu každému z nich a v okamžiku, kdy zjistí nějaké odchylky, musí zajistit nápravu. Každý podproces, ve kterém se objevují výrazné odchylky od normálu je kandidátem na vylepšení. Základní cíle managementu procesů je především snížit variabilitu uvnitř procesu na minimum. Toho lze dosáhnout tak, že procesy budou kontrolovány. Jedině kontrolované procesy jsou stabilní a stabilní procesy umožňují předvídat výsledky, což je obecným cílem.

Role managementu softwarových procesů také spočívá v provádění následujících čtyř kroků:

- definování procesů
- měření procesů
- kontrolování procesů
- zlepšování procesů

Toto jsou čtyři základní kroky, kterými by se měly řídit všechny procesy, které chceme zlepšovat. Celou situaci zachycuje Obrázek 2-2 Čtyři klíčové oblasti managementu procesů. Nyní se zmíníme velmi stručně ke každému z nich. Podrobnému popisu pak budou věnovány další kapitoly.



Obrázek 2-2 Čtyři klíčové oblasti managementu procesů (převzato z [FLO99])

Definování procesů zahrnuje činnosti spojené především s vytvářením vhodného prostředí pro provádění procesů. Jedná se o:

1. navrhování procesů
2. stanovování cílů a metrik jednotlivých procesů
3. poskytování podpory pro daný proces (lidé, materiály, vybavení, atd.)
4. zajišťování plnění daných úkolů (finance, nástroje, vzdělání, školení, atd.)

Vykonávání procesů, jak je uvedeno na obrázku oranžovou barvou, není součástí managementu procesů, ale projektového managementu. Někdy se však tyto pozice nedají dobře odlišit a často splývají. Tento krok má však stejný význam jako všechny ostatní a proto se mu bude věnovat v samostatné části této práce.

Měření procesů má jednoduchý cíl. Poskytnout co možná největší množství užitečných dat, která souvisejí s daným procesem a která je možné následně použít pro vyhodnocení kvality a výkonnosti daného procesu a jeho případného následného vylepšení. Měření procesů je také důležité z dlouhodobého pohledu. Na základě historických dat lze totiž daleko snadněji provádět aktuální rozhodnutí a lépe předvídat výsledky. Obecně lze tedy do měření procesů zahrnout tyto činnosti:

1. sběr dat o každém procesu
2. analyzování výkonnosti procesů
3. ukládání data pro pozdější potřeby (lepší predikce ceny, výkonnosti, atd.)

Hlavním úkolem kontrolování procesů je zajistit jejich chování v předem stanovených mezích. Mezi to se počítá:

1. identifikování procesů, které se vymykají předpokládanému chování
2. identifikování příčiny tohoto abnormálního chování
3. odstranění těchto příčin

Zlepšování procesů je oblastí, které se tato práce věnuje. Proces může být měřen a může být i pod kontrolou, což znamená, že je prováděn podle předem stanovených předpokladů. To ovšem neznamená, že se nedá zlepšit. K tomu je nutné celý proces nebo jeho jednotlivé podprocesy:

1. důkladně analyzovat a porozumět faktorům, které mají vliv na jejich výkonnost a stabilitu
2. naplánovat a provést kroky, které povedou k úpravě, náhradě nebo odstranění celého procesu nebo jeho částí s cílem zvýšit celkovou výkonnost a stabilitu při splnění obchodních cílů dané organizace
3. zhodnotit výsledky a náklady spojené s danou úpravou

Tolik tedy k základnímu schématu managementu procesů. Jednotlivým krokům budou věnovány další kapitoly.

Výsledkem každého procesu je obecně produkt nebo služba. Proto lze proces měřit a to z několika základních pohledů. Lze sledovat cenu, kvalitu a kvantitu produktu a také čas a úsilí potřebné k jeho vytvoření. S procesy je pak spojena řada pojmů, které je nutné znát pro lepší orientaci. Význam některých z nich je vysvětlen dále.

Výkonnost (*performance*) – jak lze očekávat, u procesů lze obecně stanovovat jejich výkonnost a to především na základě času, nákladů a kvality produktů, které jsou výsledkem daného procesu. Pokud výkonnost daného procesu klesá pod stanovené limity, pak proces není pod kontrolou. Jinak lze také říct, že proces není stabilní.

Stabilita (*stability*) – stabilita napovídá o tom, zda je proces předvídatelný. Základním krokem k dosažení stability je analýza chování procesu v čase, například pomocí grafů. Z nich lze snadno identifikovat problémy, často také určit příčiny a provést nápravná opatření.

Plnění (*compliance*) – stabilita procesu ještě nezajišťuje jeho bezchybné provádění a plnění. Je nutné si klást další otázky, jako například:

1. Je proces plně podporován tak, aby plnil cíle, ke kterým byl navržen?
2. Je proces prováděn zodpovědně?
3. Je proces v daném prostředí schopen plně fungovat?

Vypělost (*capability*) – stabilní proces, který je vykonáván zcela bez chyb, může být dále vylepšován a zdokonalován. Může se posouvat jeho tzv. vypělost, což znamená především snižování všech pozorovatelných odchylek tak, aby splňovaly předem stanovené cíle. Jinými slovy, vypělost procesu je úzce spjata s tím, jak daný proces nebo produkt dokáže plnit specifikace, se kterými byl navržen.

Vylepšení (*improvement*) – s vylepšováním procesů se pojí opět několik otázek:

1. Jaká je výkonnost daného procesu?
2. Jak snížit odchylky?
3. Jak si být jist tím, že provedené změny měly patřičný efekt?

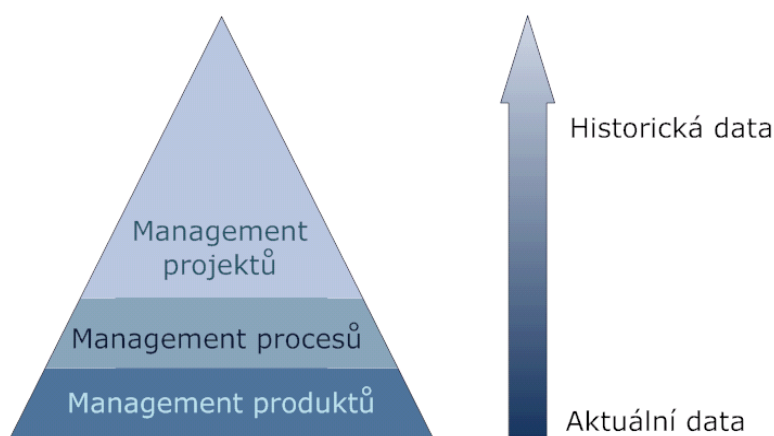
Je tedy zřejmé, že vylepšování procesů zahrnuje všechny výše uvedené kroky, tedy zvyšování výkonnosti procesu, udržování jeho stability a tím i jeho předvídatelnosti, zajištění jeho důsledného vykonávání a samozřejmě i zvyšování nároků kladených na produkt nebo proces samotný, čímž posouváme jeho vypělost.

Z hlediska dlouhodobých cílů, které procesy sledují, lze definovat tři úrovně managementu podle toho, jaké informace poskytnuté metrikami používají pro rozhodování. Jedná se o rozdělení funkcí do tříd:

1. Management projektů – zohledňuje kvalitu, cenu, funkčnost a čas potřebný k dodání výsledného produktu nebo služby. Z tohoto pohledu jsou na této úrovni rozhodování důležitá zejména historická data z předešlých projektů, podle kterých se vytvářejí splnitelné cíle.

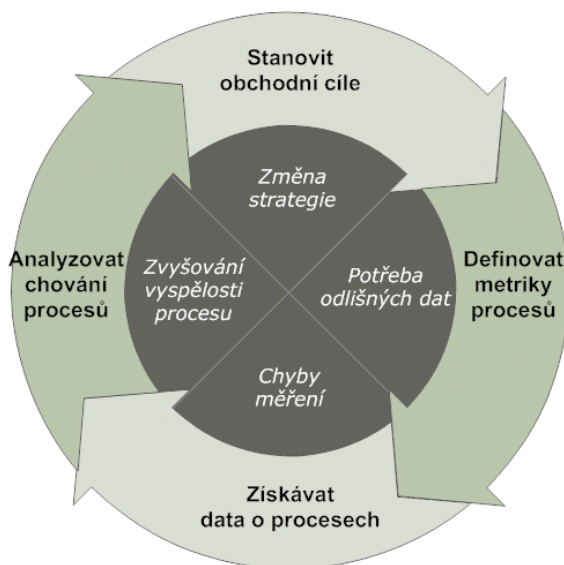
2. Management procesů – sleduje to, zda jsou procesy v dané organizaci prováděny tak, jak bylo naplánováno, zda jsou dodržovány a případně činí nutná vylepšení pro dosažení maximální kvality výsledku. Pro management procesů jsou důležitá zejména aktuální data poskytovaná procesy a případně i historická data, která pomáhají se vyvarovat chybám při změnách procesů v souvislosti z jejich změnami.
3. Management produktů – jedná se o inženýrskou úroveň, na které dochází k provádění procesů a sběru naměřených dat k jednotlivým procesům. Vedoucí vývojáři a analytici provádějí denně mnoho zásadních rozhodnutí, která jsou součástí přirozeného procesu vývoje produktu a snaží se tak zajistit maximální spokojenost zákazníků. K těmto rozhodnutím potřebují mít aktuální naměřená data, znát například výkonnost jednotlivých procesů a to, zda je vývoj v předstihu nebo je vůči plánu opožděn a tomu přizpůsobovat další kroky.

Z výše uvedeného je zřejmé, že jednotlivé úrovně managementu potřebují pro svá rozhodnutí různá data a informace s různou úrovní abstrakce. Tuto situaci zachycuje Obrázek 2-3.



Obrázek 2-3 Jednotlivé úrovně managementu

V souvislosti s měřením a zlepšováním procesů existuje základní schéma, kterým se tato činnost řídí. Jedná se o posloupnost kroků, které jsou úzce propojeny a vzájemně na sebe navazují. Obsahuje celkem čtyři fáze, které zachycuje Obrázek 2-4 Schéma měření a vylepšování procesů. Šedý kruh uprostřed obsahuje události, které mohou vyvolat potřebu vrátit se znovu k dané fázi.



Obrázek 2-4 Schéma měření a vylepšování procesů

První fází vylepšování procesů je inicializace. Začíná stanovením cílů, kterých chceme dosáhnout. Pokud budeme hovořit o vývoji software, pak budeme mít na mysli zejména o vyjasnění obchodních plánů a obchodní strategie, zhodnocení stavu trhu a posouzení konkurenceschopnosti, posouzení požadavků zákazníka na produkt nebo službu a další. V této fázi jsou jednotlivé cíle seřazeny podle priorit a jsou posouzeny možnosti jejich splnění. Mohou být pochopitelně zvoleny cíle, které jsou naprosto nové a dosud není možné posoudit jejich vlastnosti a náročnost splnění. Tyto jsou vhodnými kandidáty pro bližší zkoumání a měření.

Následuje fáze výběru a definování metrik. Podle dříve vybraných cílů jsou specifikovány metriky, které budou popisovat jednotlivé procesy. Je nezbytné stanovit způsob měření dat a jejich jednoznačný význam tak, aby nedocházelo v budoucnu k nesrovnalostem a omylům, které by mohly výsledky měření zkreslit nebo naprosto znehodnotit.

Ve třetí fázi sběru dat totiž do hry vstupuje lidský faktor. Jedná se o kritickou fázi z pohledu množství a přesnosti nasbíraných dat. Může zde dojít i přes detailně specifikovaný proces měření k mnoha problémům, které mohou mít zásadní vliv na následující fázi analyzování procesu. Této problematice bude věnována samostatná podkapitola.

Poslední je fáze analýzy získaných dat a posouzení chování procesů z hlediska jeho výkonnosti, spolehlivosti, plnění stanovených cílů a možnosti jeho dalšího vylepšení. V této fázi jsou používány především grafy, které přehledně ilustrují chování procesu. Pokud je proces vyhodnocen jako nevyhovující, je nutné identifikovat a analyzovat příčiny tohoto stavu a pokusit se o jejich odstranění nebo proces upravit. Této problematice bude věnována samostatná kapitola.

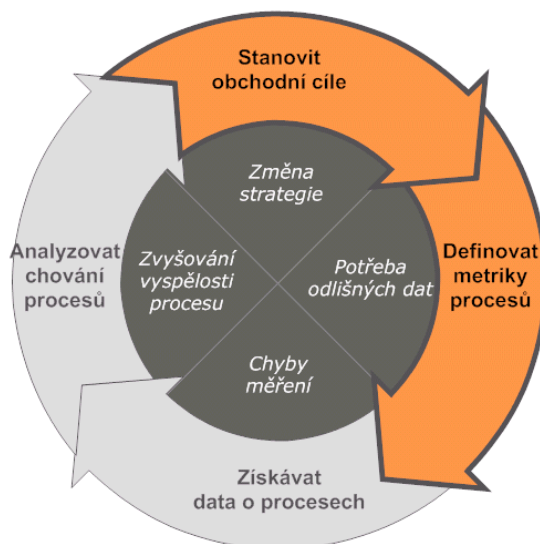
Je nutné si uvědomit, že všechny čtyři fáze na sebe navzájem navazují a jakákoliv nepřesnost nebo dokonce chyba ve kterémkoliv kroku negativně ovlivní všechny následující. Čím více chyb v brzkých fázích vývoje procesu, tím horší mohou být následky.

Poslední dvě fáze jsou prováděny opakovaně. Aktuálně získaná data jsou vždy analyzována s cílem posouzení procesu. Jedná se tedy o opakovanou činnost, která může být přerušena v případě výskytu nějaké vnější události. Pokud se například změní obchodní strategie společnosti nebo podmínky vývoje software, je nutné přehodnotit cíle a případně znovu projít příslušné fáze vylepšování procesů, tj. definovat metriky, získat data o procesech a ta následně analyzovat. Během měření dat nebo analýzy chování procesu může vzniknout potřeba nových, dosud neměřených a nespecifikovaných dat, která budou daný proces detailněji popisovat. Toto vyvolá potřebu definovat nové metriky a na jejich základě získat potřebné údaje. Výjimkou nejsou ani chyby v měření, které se, jak již bylo zmíněno výše, mohou působením lidského faktoru vyskytnout. Takto pořízená data silně ovlivňují výsledek analýzy procesu v poslední fázi, což je pochopitelně nežádoucí. Ta přichází ke slovu také v okamžiku, kdy zvažujeme možnost vylepšení procesu. Pro tento krok je nutné detailně vyhodnotit všechna získaná data a pokusit se identifikovat slabá místa v daném procesu.

Na závěr jedna důležitá poznámka ohledně managementu procesů. I přes některé názory, podle kterých je proces vývoje software vysoce intelektuální záležitostí, spočívající především na kvalitě, zkušenostech a intuici lidí podílejících se na něm, je použití metrik v současné době aktuálním tématem. Mohlo by se zdát, že softwarový proces nelze měřit, ale je to proces jako každý jiný a tak k němu lze přistupovat. Skládá se z přesně definovaných kroků a činností, které vykonávají lidé, avšak stále se jedná o činnost, která má svá pravidla a řád a je založená na dřívějších zkušenostech a pozorování. Nejedná se tedy o čistě lidskou činnost a takto je nutné na softwarový proces pohlížet. Pouze v případě, kdy jej budeme moci sledovat a vyhodnocovat jeho vlastnosti, jej budeme moci také vylepšovat, což je hlavním cílem. Podobně i definování cílů, metrik, sběr dat a jejich analýza jsou založeny na jasně definovaných a praxí odzkoušených principech, které mají svůj řád a z tohoto důvodu je nasazení metrik do oblasti zkvalitňování procesů tím nejdůležitějším krokem, kterého se není nutné obávat.

2.1 Zavádění metrik do procesů

Tato podkapitola blíže popisuje první dvě fáze schématu měření a vylepšování procesů (Obrázek 2-5 Obsah kapitoly Zavádění metrik do procesů). V první části je zdůrazněn význam a podstata stanovení obchodních cílů a jejich dopad na celý následující proces. Druhá část se zaměřuje na to, jakým způsobem vybírat a definovat metriky procesů. Jelikož se jedná o velmi náročnou fázi, ve které neexistuje návod, jak přesně postupovat, uvádí časté příčiny neúspěchů a možné zdroje chyb. Na závěr je uvedeno shrnutí všech dosud získaných znalostí.



Obrázek 2-5 Obsah kapitoly Zavádění metrik do procesů

2.2 Stanovení obchodních cílů

Tato podkapitola se zaměřuje na první fázi zdokonalování procesů. Popisuje kroky, které jsou vykonávány před samotným definováním metrik a jejich nasazením do praxe.

Procesy jsou vytvářeny za účelem splnění konkrétních cílů. Abychom mohli definovat metriky, musíme nejdříve znát procesy a ty vytváříme na základě předem stanovených cílů. Může se jednat o cíle obchodní, které souvisejí s podnikáním (například zvýšení zisků, snížení nákladů) nebo cíle související s produktem nebo službou, kterou chceme poskytovat (snížení ceny produktu, zvýšení kvality, zkrácení doby vývoje a další). Častými cíly bývá také zvýšení zajištění spokojenosti zákazníka a to zejména z pohledu dodržení dohodnutých termínů.

Všechny tyto cíle obecně určují směr, kterým se budou procesy ubírat a ovlivňuje tak zásadním způsobem všechny následující fáze měření a zkvalitňování procesu vývoje software. Během celého životního cyklu procesu je nutné všechny cíle pečlivě sledovat a případně provést nápravné kroky, které zajistí jejich bezproblémové dosažení podle předem specifikovaných kritérií.

Splnitelnost cílů je hlavním předpokladem. Při jejich vytváření musíme proto ověřit, zda je daný cíl opravdu splnitelný. K tomu je zapotřebí mít odpovídající množství lidských zdrojů, materiálu, vybavení, prostředků a také financí. Pro dosažení cílů je potřebný také dostatečný čas. Většina procesů totiž v časové tísní vykazuje nižší výkonnost a častější jsou také projevy nestability.

Stanovení cílů je velmi úzce spjato s definováním procesů a následným stanovováním metrik. Tato koncepce se nazývá *GQM - Goal Question Metrics*. Byla vytvořena v roce 1984 autory Basili a Weiss jako výsledek praktického a akademického výzkumu. Jedná se o koncept, který popisuje, jak stanovit splnitelné cíle, z nich odvodit otázky, které následně pomohou definovat metriky. Zásadním

je právě vytváření otázek, které by se měly zaměřovat především na způsob, jakým lze dané cíle dosáhnout a jak tyto cíle souvisejí se zdokonalováním procesů.

1. Stanovit cíle, kterých chceme dosáhnout. Může se jednat i o cíl zlepšování nebo zvyšování kvality procesů.
2. Položit otázky, jakým způsobem bude dosaženo splnění daných cílů nebo jakým způsobem se splnění cílů projeví.
3. Definovat metriky, které budou indikovat splnění cílů.

Tato kapitola popisovala první fázi procesu měření a zvyšování kvality software. V souvislosti s tím zmínila metodiku *GQM (Goal Question Metrics)*, která se často používá při stanovování cílů.

2.3 Definice metrik procesů

Následující podkapitola popisuje postup, jak vhodným způsobem definovat metriky procesů a ty následně zavést do praxe. Jedná se o druhý krok schématu zvyšování kvality procesů.

Definování metrik procesů má tři základní fáze, které na sebe navzájem navazují. Následující fáze vždy staví na informacích z fáze jí předcházející a proto nemůže být žádná z nich vynechána. Tvoří tak pevnou posloupnost kroků, které ukazuje Obrázek 2-6 Fáze definování metrik. Následující podkapitoly se budou zvláště věnovat jednotlivým krokům.



Obrázek 2-6 Fáze definování metrik

2.3.1 Identifikace rizik

V této podkapitole se zaměříme na postup, jak efektivně identifikovat rizika, která jsou spojena s procesy, což představuje první krok definice metrik.

Všechny procesy bez výjimek s sebou nesou jistou míru rizika. Vždy je však velmi důležité na počátku definování metrik odhalit pokud možno co nejvíce rizik, která mohou před, během nebo po provedení procesu nastat a zhodnotit také jejich význam a případný dopad.

Pokud tedy známe cíle, kterých chceme dosáhnout, a procesy, které tyto cíle sledují, přistoupíme k analýze rizik. Nejdříve je nutné se zaměřit na ty procesy, které v minulosti vykazovaly nějaké anomálie nebo nízkou výkonnost. Postupně vytváříme seznam potenciálně rizikových procesů, do kterého dále přidáváme ty, které byly v minulosti nestabilní a jejich chování by tudíž mohlo být hůře předvídatelné. Je nutné také věnovat pozornost novým technologiím, které mohou procesy do sebe začleňovat. Toto opět s sebou nese jistou míru neurčitosti, která může způsobit problémy.

Z tohoto seznamu potenciálně rizikových procesů se postupně zaměříme na každý zvláště a zkoumáme jeho kritické body. To znamená, že hledáme možné příčiny neúspěchu procesu a to z různých úhlů pohledu. Jedná se hlavně o schopnost splnit stanovené cíle v daném čase, za rozumnou cenu a v přijatelné kvalitě. Z obchodního hlediska tedy hledáme na kritických procesech vlastnosti, které mohou vést obecně k nespokojenosti zákazníka. Tyto atributy lze nalézt pomocí různých metod. V praxi se nejčastěji používají dva kroky:

1. identifikovat kritické body pomocí atributů procesů
2. pomocí otázek ověřit platnost kritických bodů, případně některé z nich doplnit

První krok, identifikace kritických bodů procesu, vychází z jeho definice, která je uvedena v předešlé kapitole. Ta uvažuje o procesu jako o množině navzájem souvisejících prvků, mezi které počítáme především lidské zdroje, materiál, vybavení a činnosti. Dále se mezi ně zahrnují všechny vstupy a výstupy procesu. Jednotlivé atributy zkoumáme a hledáme kritické body a možná rizika, která s nimi souvisejí. Mohou nám pomoci různé grafy, diagramy, náčrtky procesů a další vizuální pomůcky. Pokud tedy budeme uvažovat například o lidských zdrojích, je vhodné se ptát po kvalitě a zkušenostech všech pracovníků, jejich zodpovědnosti a přístupu k práci nebo po tom, jak se chovají v náročných situacích nebo když jsou pod psychickým tlakem.

Druhý krok ověřuje platnost námi získaných kritických bodů. Pokládáme takové otázky, které platnost jednotlivých rizik buď potvrdí nebo vyvrátí. Ptáme se tedy například:

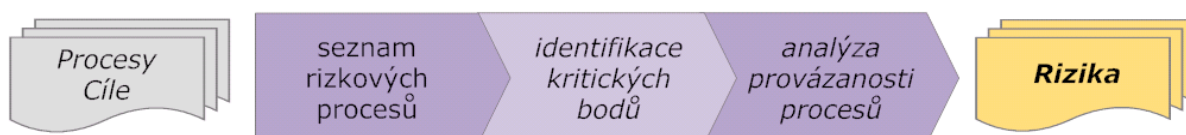
- Co určuje kvalitu produktu?
- Co zákazník požaduje?
- Co může být příčinou neúspěchu?
- Co nemůžeme kontrolovat nebo ovlivnit?
- Jaké jsou naše výkonnostní limity?

Může také nastat situace, kdy nalezneme nové kritické body, o kterých jsme dosud neuvažovali, což je v této fázi žádoucí. S rostoucí složitostí procesu roste také počet otázek, které je možné pokládat. V takovýchto situacích může být vhodné použít strukturované diagramy popisující tok činností daným procesem, které pomohou se v procesu lépe zorientovat.

Pro doplnění je dobré poznamenat, že některé kritické body jsou pro všechny procesy společné a jsou nejčastěji spojeny s otázkami kvality a ceny výsledného produktu. Například počet chyb je limitujícím faktorem mnoha procesů a má mnoho negativních dopadů. Chyby vyžadují úsilí a finance na jejich odstranění mohou vést k nespokojenosti zákazníka. U všech procesů je tedy nutné se zaměřit na všechny potenciální zdroje chyb a pozorně je analyzovat. Jedním ze společných kritických bodů je i čas. Ten hraje významnou roli ve všech fázích procesu. Při nedodržení termínů předání produktu zákazníkovi nebo i při překračování stanoveného plánu softwarového procesu vzniká velké množství chyb. Lze také zkoumat dodržování stanoveného plánu z pohledu množství a kvality práce v daném čase. Toto je další limitující bod, který je společný pro všechny procesy, stejně jako celková cena produktu. Všechny výše uvedené atributy lze do jisté míry ovlivnit nebo úplně odstranit pečlivou prací, používáním nových a kvalitnějších technologií, řádným školením nebo úpravou procesů, což zahrnuje i odstranění nepotřebných kroků.

Dosud byla pozornost věnována pouze samostatným procesům. Tento pohled však není zcela dostačující. Je nutné uvážit vzájemné vztahy všech procesů a jejich podprocesů, které mohou být složité propojeny. Někdy mohou na sebe procesy navazovat, někdy může ukončení jednoho procesu podmiňovat spuštění druhého, což se může snadno stát zdrojem dalších rizikových faktorů. K analýze provázanosti procesů je pochopitelně nutná jejich důkladná znalost. Provádí se nejčastěji pomocí diagramů a nákresů a jejím výsledkem je opět seznam rizik.

Celý proces identifikace rizik je na Obrázek 2-7.



Obrázek 2-7 Proces identifikace rizik

Celý proces identifikace rizik je prvním krokem při zavádění metrik do procesů. Skládá se ze tří kroků. Nejdříve vytvoříme v prvním kroku seznam všech rizikových procesů. Z něj postupně analyzujeme každý proces zvlášť a identifikujeme jeho kritické body pomocí zkoumání jeho atributů.

Posledním krokem je analýza provázanosti procesů. Jedná se zde o zhodnocení rizik, která plynou ze vzájemné provázanosti procesů mezi sebou. V těchto třech krocích jsou stanovena rizika, která s prováděním procesů souvisejí a která v další fázi pomáhají vybrat konkrétní metriky, které budou tato rizika sledovat a poskytnou tak dlouhodobou kontrolu nad celým procesem a jeho stabilitou. Umožní zároveň jeho lepší předvídatelnost a vylepšování.

2.3.2 Výběr a definice metrik

Nyní zaměříme pozornost na výběr metrik a jejich přesnou definici. Tato podkapitola zahrnuje druhý krok procesu definování metrik, který následuje po identifikaci rizik spojených s procesy. Budou zde popsány dvě věci, které je nutné provést před zavedením metrik do praxe.

Výběr a definice metrik jsou dvě naprosto odlišné činnosti, které však spolu velice úzce souvisejí. Jejich význam spočívá v tom, že se snažíme k danému procesu vybrat vhodné metriky, které před jejich nasazením do praxe musíme nejdříve řádně definovat, čímž zajistíme v budoucnosti jednoznačnost celého procesu sběru a interpretace dat. Jinými slovy lze říct, že vytváříme funkční definice metrik, které umožňují kontrolu a zvyšování kvality procesů.

Nejdříve se zaměříme na výběr vhodných metrik. Proces se skládá z několika navzájem propojených atributů (viz. definice procesu), které se pochopitelně dají popsat odpovídajícími veličinami. Tohle je základní princip metrik, které popisují procesy. Můžeme však také měřit atributy výsledných produktů a služeb. Rozdíl je v charakteru a účelu získaných dat. Metriky lze tedy rozdělit do dvou kategorií:

1. metriky atributů procesů (čas, náklady, úsilí, použité zdroje, počet chyb v procesu, atd.)
2. metriky produktů nebo služeb, které jsou výsledkem procesu (funkčnost, velikost, kvalita z pohledu zákazníka nebo koncového uživatele, atd.)

Pro úplnost je nutné zmínit, že toto rozdělení je pouze orientační. V procesech lze zkoumat a měřit také atributy použitých zdrojů. Konkrétním příkladem takovéto metriky může být čas procesoru a množství operační paměti spotřebované konkrétní aplikací.

Všechny metriky poskytují určitá data, která lze použít při analýze chování procesu nebo k určení jeho výkonnosti. Lze je také použít pro predikci budoucího chování a určení míry stability a vyspělosti procesu. Jejich výběr však nemusí být vůbec jednoduchý. Snahou je především určit takové metriky, které popisují atributy procesů a atributy produktů s ohledem na další rozhodování. Jedná se proto o zcela zásadní krok, který má vliv na všechny následující fáze zvyšování kvality procesu. Neexistuje však žádné pravidlo pro výběr vhodných metrik. Lze se řídit pouze následujícími kritérii, která by měly všechny metriky splňovat:

- musejí úzce souviset s kritickými body, které chceme zkoumat (čas, cena, kvalita)
- musejí mít vysokou informační hodnotu (přinášet data, která vypovídají co nejvíce o produktu a která jsou silně závislá na mnoha jeho attributech)
- musejí být ověřeny (musejí splňovat účel, se kterým byly navrženy)
- musejí poskytovat konzistentní a dobře definovaná data
- jejich sběr musí být snadný a nesmí vyžadovat zvýšené úsilí
- musejí vhodným způsobem odrážet nestabilitu a odchylky v procesu
- musejí poskytovat nejen informace o chybách, ale také o jejich možných příčinách

Metriky se tedy snažíme vybírat takovým způsobem, aby co nejvíce reflektovaly kvalitu výsledku procesu s ohledem na cíle, kterých chceme dosáhnout. Snahou je volit takové atributy, které jsou dominantní nebo jiným způsobem vhodně charakterizují proces jako celek. Nezbytnou součástí je také ověření toho, zda námi vybrané metriky opravdu podporují to, co jsme si stanovili jako cíl měření. Je vhodné si načrtnout různé grafy, které ukazují data získaná z metrik a posoudit jejich

význam a použitelnost. Někdy mohou totiž být výsledky měření, které na počátku při výběru metrik považujeme za přínosné, naprosto nevhodné. Může se také stát, že několik zdánlivě nesouvisejících metrik přináší stejné výsledky, pouze v jiné formě. Proto data, mezi kterými se vyskytuje silná korelace, je nutné odstranit, protože v opačném případě by pouze zvyšovaly úsilí k jejich získání a nepřinesly by nic nového. Mezi několika stejnými metrikami vybíráme takovou, která splňuje co nejvíce výše uvedených bodů.

Pro přesnější představu následuje výčet nejdůležitějších metrik použitelných pro popis atributů procesu a jeho výsledků.

Metriky procesu

Lidské zdroje, čas, cena, energie, materiál, školení, náklady na vývoj týmu, počet přesčasových hodin, procento splněných milníků, počet požadavků zákazníka, počet změněných požadavků, počet nedořešených požadavků, měření fáze analýzy, návrhu, implementace a testování, spokojenost zákazníka, počet změn v návrhu, počet iterací, počet člověkohodin na tisíc řádek kódu, procento testovaných požadavků, plnění stanovených plánů, počet opakování procesu, kvalita výstupního produktu, peněžní zisk.

Metriky produktu

Počet požadavků, kvalita, cena, čas, návrh, dokumentace, množství chyb, počet chyb na tisíc řádek kódu, střední doba mezi poruchami, výsledky testů, velikost, stabilita, rozpočet, paměťové nároky, nároky na procesor, zpoždění proti plánu, počet hodin strávených vývojem, testováním, opravou chyb a dokumentováním, počet řádek kódu, počet spustitelných řádek kódu, počet řádek dokumentace v kódu, počet modulů, počet objektů, datová propustnost, počet transakcí za jednotku času, škálovatelnost, udržitelnost, opakovanou použitelnost, jednoduchost použití a mnohé další.

Právě jsme si ukázali postup, jak vybrat vhodné metriky a nyní se zaměříme na to, jak tyto metriky správně definovat. Cílem tohoto kroku je zajistit jednoznačnou a nezaměnitelnou interpretaci metrik v rámci daného procesu a všech jeho částí. Je nutné také specifikovat postup, jakým jsou všechna data získávána. Dále se zaměříme na pojem *funkční definice metriky* procesu, který s výše uvedenými požadavky souvisí.

Nejdříve podrobně o tom, jaký význam definice metriky má. Poté, co jsme našli vyhovující metriky, je nutné je vhodně definovat. Uvedení pouze jména metriky nestačí. Je nutné ujasnit a popsat jak přesně je daná metrika získávána, aby bylo možné zaznamenaná data jednoznačně interpretovat. Definice metrik ovšem s sebou nese i jistá rizika. Nejběžnější je záměna použití metrik k jiným účelům, než byly původně vytvořeny. Takovou metrikou může být například úsilí vynaložené na proces. Každodenně zaznamenávaná data pomáhají určit míru náročnosti jednotlivých fází procesu, ovšem mohou být také snadno zaměněna za nástroj, kterým se měří výkonnost jednotlivých zaměstnanců, což ovšem není původní účel. Takovýchto příkladů bychom našli mnoho. Proto je vhodné na tomto místě uvést několik základních problémů, které při definici metrik mohou nastat.

- **Různé úrovně rozhodování používají různá data.** Tento fakt byl již diskutován v předešlé kapitole. Jednoduše lze konstatovat, že data pořízená s jistým účelem by neměla být používána jinak. V opačném případě může dojít k různým nekonzistencím při interpretaci těchto dat a při jejich analýze.
- **Každý proces používá specifický způsob sběru dat.** Tyto postupy není nezbytně nutné měnit jen proto, abychom vyhověli novým požadavkům. Naopak by mělo být snahou sběr dat co nejvíce ulehčit a pokud možno i automatizovat. Přizpůsobovat by se měl hlavně způsob následné analýzy a vyhodnocování dat.

- **Zajistit jednoznačnou definici a výklad všech metrik je velmi náročné.** Tento požadavek je důležitý především v dynamických a rychle se měnících prostředích nebo pokud metriky používá více subjektů. Je nezbytně nutné, aby byly metriky vykládány jednotným způsobem a nedocházelo k nepřesnostem nebo omylům.
- **Je náročné vytvořit slovník pojmů pro komunikaci metrik.** Společně s měřením procesů přichází i nutnost získaná data a znalosti předávat dále. K tomuto účelu je nezbytné komunikovat za pomoci různých nástrojů, ať už se jedná o grafy, tabulky, obrázky a nebo slovní popisy. Vytvoření formálních metod komunikace je velmi náročné a v mnohých případech i nemožné. Z tohoto důvodu mohou při předávání informací mezi subjekty vznikat mnohá nedorozumění, která mohou vést k nepřesnostem v celém procesu.

Jak je vidět z předchozích bodů, vytvořit jednoznačnou definici metrik je velmi náročný a zodpovědný proces. S tímto souvisí pojem tzv. „*funkční definice*“. Jedná se o správným způsobem definovanou metriku, která zohledňuje všechny výše uvedené problémy a která dále splňuje následující kritéria:

- **Jednoznačnost** – zaručuje přesnou a neměnnou interpretaci dat, která metrika specifikuje. Dále tento požadavek musí zajistit, aby všechny subjekty, které budou s daty pracovat, přesně věděly, co výsledky v sobě zahrnují a co naopak neobsahují.
- **Opakovatelnost** – je to schopnost opakovat dané měření. Důležitá je tato podmínka především pokud je proces s danou metrikou nasazen v jiném prostředí a data jsou pořizována jinými lidmi.
- **Dohledatelnost** – jedná se o rozšíření metrik nad rámec jejich definice. Metriky by měly umožňovat nejenom data pořizovat, ale umět je také zpětně dohledat. Přesněji určit jejich původ z hlediska času, jejich zdroje, stavu procesu, prostředí a dalších vlastností. Tento požadavek je významný především v situaci zvyšování kvality procesu a také v okamžiku, kdy je proces nestabilní. Díky tomuto kritériu je totiž snadné najít zdroj nechtěných stavů a podniknout kroky k nápravě.

S otázkou jednoznačnosti souvisí několik otázek, které bychom měli při jejím posuzování být schopni zodpovědět. Především je nutné si uvědomit, co opravdu daná metrika poskytuje a co získaná čísla znamenají. Potom je dobré znát způsob, jakým jsou data získávána. Dále co je jejich zdrojem a kdo tato data zaznamenává, jak často a za jakých podmínek. Více o této problematice v následující podkapitole. V poslední řadě bychom měli mít představu o tom, jak by měla získaná data vypadat. Jinými slovy, dopředu stanovit požadavky a omezení na hodnoty těchto dat.

Na závěr této podkapitoly malé shrnutí. Stanovení metrik procesů zahrnuje dva kroky. Prvním je výběr vhodných metrik. Můžeme měřit atributy procesů nebo atributy produktů. Druhým krokem je definice metrik a s ní související pojem *funkční definice*, který v sobě zahrnuje tři požadavky na jednoznačnost, opakovatelnost a dohledatelnost všech dat. Poslední bod úzce souvisí se zvyšováním kvality procesu.

2.3.3 Začlenění metrik do procesu

Obsahem této podkapitoly je popsat činnost, během které dochází k integraci metrik do procesů.

Účelem této fáze je co nejvhodnějším a nejefektivnějším způsobem začlenit definované metriky do samotného procesu. Toto zahrnuje několik kroků, které zajistí celkovou funkčnost měření. Je nutné především zhodnotit aktuální stav zaznamenat všechny již používané metriky. V dalším kroku dojde k posouzení toho, zda některé z nich již nesplňují kritéria, která jsme stanovili a neposkytují data, která chceme získávat. Následuje poslední a nejdůležitější krok, kdy přesně stanovíme proces sběru dat a ten následně uvedeme do praxe. Nyní k jednotlivým krokům podrobněji.

V první fázi nasazování metrik do procesů musí dojít k analýze současného stavu. Vždy je totiž lepší budovat systém metrik na již existujících a pokud možno osvědčených základech. V opačném případě může být proces začleňování metrik obtížnější a dokonce se může setkat s odporem. Stejně tak tomu je i v případě, kdy se snažíme zásadním způsobem pozměnit nebo zcela nahradit celý proces měření.

Pokud již některé metriky existují a jsou používány, musí dojít k jejich analýze. Jsou posouzeny z hlediska vhodnosti a toho, zda splňují kritéria tzv. funkčních metrik (více v předešlé podkapitole). Tyto metriky jsou pak použity v procesu měření a není nutné je měnit. V opačném případě, kdy metriky zcela nevyhovují představám a cílům, je vhodné zejména:

- upravit data tak, aby vyhovovala potřebám
- upravit potřeby tak, aby vyhovovaly získávaným datům
- doplnit chybějící atributy

Posledním a nejdůležitějším krokem je převedení všech dosavadních výsledků do praxe. Toto zahrnuje především vytvoření specifikací a plánů měření, které budou definovat způsob, jakým data budou získávána a stanovení zodpovědností za provádění jednotlivých kroků a zajištění jejich plnění.

Při vytváření plánu měření je nutné zvážit a zdokumentovat:

- formát a význam dat, která budou metriky poskytovat
- rozsah a omezení hodnot získávaných dat
- frekvenci získávání dat
- podpůrné prostředky sloužící při sběru dat (aplikace, nástroje, formuláře a další)
- způsob ukládání dat
- systém přístupových práv k dané databázi naměřených dat
- způsob použití dat
- způsob analýzy dat

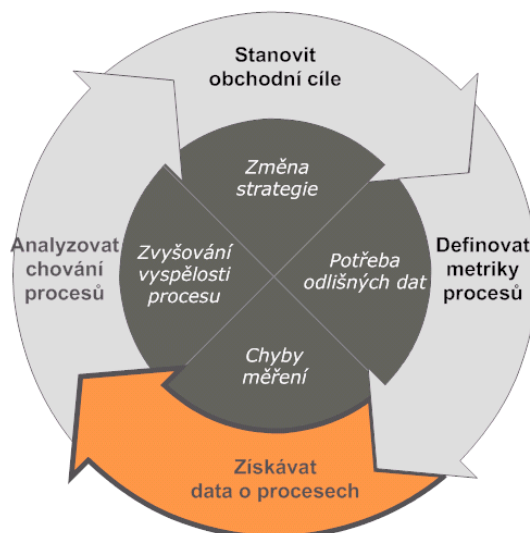
Na tomto místě je vhodné ještě jednou zdůraznit, že metriky by měly sloužit účelu, ke kterému byly vytvořeny. Proto je nutné při zavádění metrik vysvětlit všem zúčastněným jejich smysl a zdůraznit jim, že metriky neslouží jako prostředek k monitorování výkonnosti pracovníků. Proto musí být zejména zvážena způsob ukládání dat a zároveň musí být vytvořen systém přístupu k takovéto databázi, aby byla zajištěna ochrana dat před jejich zneužitím.

Na základě předešlých úvah je vytvořen plán měření. Ten transformuje cíle a požadavky společně se všemi omezeními na konkrétní posloupnost kroků vykonávaných během procesu sběru dat. Měl by především vytvořit jasnou představu o cílech a rozsahu měření a definovat souvislosti mezi jednotlivými kroky. Dále musí obsahovat úkoly, které budou prováděny, a přiřadit odpovědnost subjektům za jejich splnění. Musí také specifikovat všechny použité zdroje, poskytnout podporu pro sledování vývoje procesu měření a management rizik. Důležité je také definovat schéma, které zaručí bezproblémový průběh měření a zajistí přiměřený dohled nad jeho vykonáváním.

Na závěr shrnutí této podkapitoly. Začlenění metrik do procesu je poslední fází při definování metrik. Zahrnuje tři hlavní kroky – analýza současného stavu z pohledu již používaných metrik a jejich posouzení z hlediska opětovné použitelnosti, což pomůže proces zavádění metrik velmi usnadnit. Posledním krokem je vytvoření plánu měření. Ten zahrnuje zejména kroky, které budou při měření vykonávány a přiřazuje zodpovědnosti za jejich splnění jednotlivým subjektům. Důležité je také v této fázi zdůraznit význam a účel metrik.

2.4 Získávání dat o procesech

V této podkapitole zaměříme svoji pozornost na získávání a uchovávání dat poskytovaných ve fázi měření procesů. V první části se budeme věnovat procesu získávání dat, který začíná samotným měřením procesů. Budou zmíněny všechny problémy, které mohou nastat. Pozornost v souvislosti s tímto bude věnována zejména lidskému faktoru. Dále bude vyzdvihnout význam automatizace při sběru dat a vysvětleny některé pojmy související s touto problematikou. Druhá část se bude věnovat ukládání získaných dat do databází. Zaměří se především na otázku výběru vhodného systému a shrne požadavky, které jsou na něj kladeny.



Obrázek 2-8 Obsah kapitoly Získávání dat o procesech

2.4.1 Získávání dat

Tato podkapitola zahrnuje první krok při získávání dat. Uvádí základní úlohy, které s ním souvisejí a zdůrazňuje význam lidského faktoru a chyb z něj plynoucích.

Měření procesů začíná sběrem dat. V předešlých krocích jsme jednoznačně definovali metriky a ty pak pečlivě integrovali do procesu, což zahrnovalo nasazení správných lidí, nástrojů a technik na správná místa procesu. Výsledky poskytují data, která se ukládají pro pozdější analýzu chování procesů.

Základní úlohy související se sběrem dat lze popsat pomocí několika bodů:

- Vytvořit metody a získat nástroje, které budou podporovat sběr dat.
- Sestavit tým, který zajistí vykonávání této činnosti.
- Sbírat data o každém procesu, který chceme monitorovat.
- Kontrolovat dodržování postupů při získávání dat.

Úsilí vynaložené na sběr dat musí být co nejmenší. Je výhodné použít automatizované nástroje, které umějí například zjistit počet řádků kódu, vyhodnocovat automaticky různé statistiky a podávat zprávy o průběhu některých úkolů. S tím také souvisí jednoduchost, se kterou by měla být data dostupná. Získávat složitým způsobem metriky, které poskytují poměrně malé množství informací je neefektivní. Například vyhodnocování provázanosti tříd u objektově orientovaných jazyků je manuálně nesmírně zdlouhavá činnost, při které může dojít k mnoha chybám.

Je tedy vhodné použít v co největší míře automatizovaných technik sběru dat všude tam, kde je to alespoň trochu možné. Pokud jsme nuceni z jakéhokoliv důvodu získávat data ručně, musíme

zajistit jejich jednoznačnost, což je předmětem předcházející fáze definování metrik. Pouze v tomto případě zajistíme, že získaná data budou korektní a budou odpovídat jejich definici. Dosáhneme tak odstranění šumu, zpřesnění všech hodnot a lepších výsledků ve všech následujících fázích. Toto má pak také zásadní vliv na množství úsilí související s analýzou získaných dat a hraje důležitou roli při zdokonalování procesů. Nelze totiž důsledně posuzovat výkonnost a stabilitu procesu, pokud nemáme k dispozici kvalitní data.

Jelikož je proces získávání dat častou a pravidelně se opakující činností, je velmi náchylný k samovolným změnám a odchylkám od stanoveného plánu. Je to způsobeno především lidským faktorem, který zahrnuje nejen lidské chyby, ale i nedůslednost, neschopnost soustředit se na opakované činnosti a další negativní vlastnosti lidské práce. Výsledkem jsou nekvalitní, neúplná, chybějící nebo chybná data o procesech, což pochopitelně nelze akceptovat. Dále může dojít v důsledku nezodpovědnému přístupu k měření k zaokrouhlovacím chybám, které jsou jen těžko odhalitelné a zásadním způsobem tak znehodnocují celý proces měření.



Velikost obdélníku odpovídá závažnosti problému a míře úsilí, které je nutné vynaložit na jeho odstranění.

Obrázek 2-9 Vliv lidského faktoru a chyby z něj plynoucí

Dva způsoby, jakými lze učinit sběr dat lepší a odstranit z něj vliv lidského faktoru jsou:

- přiřadit zodpovědnost za provádění konkrétních měření konkrétním osobám
- zajistit dohled nad správností provádění měření a získávání dat a dodržování stanoveného plánu měření

Pokud určíme, kdo je zodpovědný za konkrétní kroky a měření, lépe pak budeme moci identifikovat příčinu nedostatků a snadněji a rychleji také dojde k jejich nápravě. Lidé musejí svoji práci vykonávat zodpovědně, čehož ovšem může být někdy velmi těžké dosáhnout. Je zde ještě jeden fakt, který byl uveden v předchozí podkapitole. Jedná se o jakýsi strach z metrik. Všem členům podílejícím se na získávání dat musí být zřejmý účel tohoto konání. Musejí přesně vědět, k čemu jsou metriky používány a za jakým účelem jsou shromažďovány a být ujištěni, že nikdy nebudou obráceny proti nim. Vhodný způsob, jak tuto myšlenku v praxi prezentovat, je na společné poradě všech členů týmu a managementu projektu a managementu procesů.

Definovat zodpovědnost ještě není krokem, který by postačoval k dokonalému zvládnutí procesu získávání metrik. Stále je nutné pravidelně, například formou mítinků, porad a diskusí, kontrolovat plnění plánu měření. Je nutné zaměřit pozornost na jakékoliv odchylky a nedostatky, které mohou vzniknout. Získaná data by měla splňovat několik základních vlastností:

1. **platnost** – je nutné ověřit, zda pořízená data odpovídají jejich specifikaci a neobsahují chyby. Typicky se hodnotí následující kritéria:
 - správnost použitého datového typu (textová data musejí být zaznamenána řetězci znaků a číselná data pochopitelně čísly)
 - správnost formátu dat (všechna data mají specifikován svůj formát, ať se jedná o datum, měnu, název, identifikační číslo, kód, priority a mnohé další)
 - rozsah hodnot (ověřuje se platnost dat z pohledu omezení plynoucích ze specifikace)
 - úplnost (data musejí zahrnovat i další informace, které například identifikují jejich zdroj, nástroje použité pro jejich získání, okolnosti při jejich získání, čas naměření, autor a další)
2. **současnost** – pod tímto pojmem se myslí pořizování dvou a více metrik současně s ohledem na čas jejich výskytu. Jedná se o to, že data by měla být pořizována v daných okamžicích (například pravidelně ve stanovených intervalech) a současně. V opačném případě velmi často nastává při vyhodnocování chování procesu k mnoha problémům. Metriky pořizované v libovolném čase a bez jakýchkoliv závislostí jsou jen málo použitelné, protože neumožňují hledat příčiny například nestability procesu. Může se také jednat o problém, kdy posuzujeme výkonnost procesu a nezohledňujeme všechny jeho vstupy, stavy a výstupy. Příkladem může být dovolená, státní svátky, nemoc, služební cesty, školení a jiné překážky v práci, kdy výkonnost měřená v pravidelných intervalech (např. jednou za 14 dní) vykazuje v tomto období nízké hodnoty.
3. **konzistence** – její posouzení je velmi náročnou činností. Je spojena s posuzováním vhodnosti a použitelnosti dat k pozdější analýze. Vyžaduje dostatečné znalosti dříve získaných dat. Zahrnuje kontrolu toho, zda jsou data opravdu získaná v uvedeném období a ve správných dnech a zajišťuje vykonání případných změn v okamžiku, kdy dojde například k personálním změnám.
4. **správnost** – schopnost ověřit, zda daná data opravdu popisují správné atributy. K tomu je nutné mít přesně definované metriky a znát jejich význam. Nesmějí existovat žádné sporné body a nezodpovězené otázky.

2.4.2 Uchovávání dat

Tato podkapitola popisuje postupy a doporučení, kterými je vhodné se řídit při výběru systému pro uchovávání dat.

Potřeba uchovávat data je nezbytnou součástí měření procesů. V současné době je používání papírových formulářů ke sběru dat nevhodné a pro následnou analýzy nepoužitelné. S výhodou se proto využívá počítačových systémů, které poskytují podporu a tvoří nástroje, které jsou integrované do procesů a slouží k získávání, ukládání a později i k analýze naměřených dat.

Často je zapotřebí jedna nebo více databází. Záleží také na době, po jakou chceme data uchovávat pro pozdější použití. V případě managementu procesů a managementu projektů, které využívají historická data pro potřeby rozhodování a předpovídání chování procesů, je nejlepší mít k dispozici data i po ukončení projektu a to i několik let. Tato kritéria a další faktory pochopitelně zásadně ovlivňují výběr systému ukládání dat. V praxi je dobré postupovat tak, že se budeme snažit využít v co největší míře již zavedených systémů. Pokud například existuje projektová databáze, může být vhodným nástrojem pro uchovávání dat. Opět je dobré poznamenat, že nasazování metrik

by se mělo celého procesu dotknout co nejméně. Pokud je to nezbytné, například z kapacitních důvod, může být pro ukládání dat nasazena zvláštní databáze. Nástroje a aplikace, které s touto databází budou spolupracovat a zaznamenávat a ukládat data od uživatelů, by měly zohledňovat několik důležitých faktů:

- získávat a ukládat data společně s jejich definicí, ne pouze naměřená data
- svázat naměřené hodnoty s okolnostmi, které vedly k jejich pořízení, a s tím, jak byly pořízeny (produkt, proces, prostředí, stav projektu, události, metoda pořízení dat, nástroje)
- svázat naměřené hodnoty s měřenými atributy
- umožňovat změny v procesech
- umožňovat měnit definici metrik a umět na takovou změnu reagovat

Databázový systém by měl také zamezit ukládání nepřímých dat. Jedná se nejčastěji o hodnoty dosažené nějakým aritmetickým výpočtem, například agregované hodnoty jako suma, průměr a další. Existují k tomu čtyři důvody:

- do databáze se vnaší redundance
- je složité zachovat platnost všech dat (při změně jedné hodnoty se mohou stát všechny nepřímé hodnoty neaktuální)
- pokud ukládáme pouze nepřímé hodnoty, můžeme snadno ztratit některé důležité informace
- v budoucnu může vzniknout potřeba vypočítat nepřímé hodnoty jiným způsobem a k tomu jsou zdrojová data nezbytně potřebná

Při zavádění databází bychom měli zvážit ještě několik dalších aspektů. V souvislosti s provozem databáze mohou vznikat konflikty při souběžném přístupu. Jedná se především o zápis nových dat a vznik možných duplicit. Tento stav je možné ošetřit jednak na aplikační úrovni pomocí dočasných zámků a nebo na úrovni administrativní pomocí důkladného rozdělení zodpovědností při definování metrik. Dále je nutné zajistit, aby jednou platná a zkontrolovaná data nebylo možné změnit a třeba i neúmyslně je tímto poškodit.

Řízení přístupu byla věnována již velká pozornost. Tento požadavek má souvislost s bezpečností a utajením citlivých dat. Je potřeba zvážit, kdo bude mít přístup k databázi, kdo bude moci vkládat a kdo měnit data, kdo bude zodpovídat za kontrolu a integritu dat a také kdo bude mít právo přidělovat jednotlivé role dalším uživatelům. Toto je otázka administrativního zabezpečení, které by mělo být nedílnou součástí každého systému pro uchovávání dat o procesech.

Bezpečnost je dalším důležitým požadavkem. Obsahuje výše uvedené řízení přístupu k databázi, ale mimo to také ochranu dat před ztrátou či odcizením, pravidelné zálohování celého systému a v neposlední řadě také zajištění bezpečnosti proti zneužití systému například pomocí hesel.

Na závěr je také podstatné zvážit výběr hardwarového a softwarového vybavení a zajistit potřebnou podporu celému systému. Jedná se o zabezpečení plynulého chodu, případné rychlé odstranění výpadků, údržbu systému, zálohování a také o zajištění řádného proškolení všech zúčastněných stran.

Na závěr shrnutí této podkapitoly. Problematiku uchovávání dat je v zásadě nutné řešit pouze na začátku při zavádění metrik do procesů. Je nutné vhodným způsobem vybrat a definovat systém pro získávání a uchovávání dat, na který je kladeno velké množství požadavků. Mezi ty nejvýznamnější patří požadavek na bezpečnost a odolnost proti zneužití a z dlouhodobého hlediska také schopnost uchovávat i historická data pro pozdější použití.

3 Analýza chování procesů

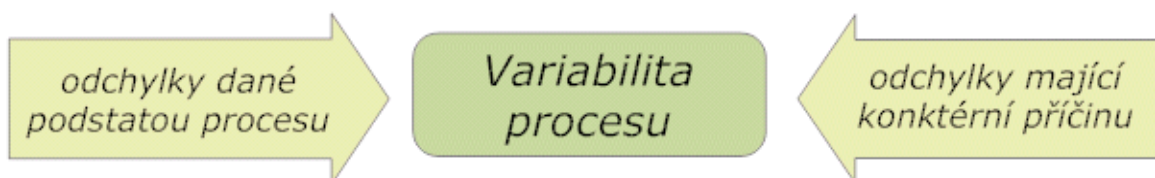
Tato kapitola přináší teoretické poznatky o analýze chování procesu. Zdůrazňuje význam stability procesu ve vztahu k jeho předvídatelnosti, popisuje příčiny variability v procesech a obsahuje také popis všech typů grafů, které jsou hlavní součástí statistického přístupu k analýze chování procesu. Analýza získaných dat spočívá v jejich transformaci na informace, které nám pomohou proces lépe pochopit, popsat nebo jej zkvalitnit. Pro tyto účely se používají různé grafy, o kterých se zmíníme dál v textu. Cílem analýzy chování procesů je identifikovat odchylky v naměřených datech, které signalizují nestabilitu procesu. Jak bylo uvedeno již dříve, pouze stabilní procesy jsou dobře předvídatelné a tomuto stavu se snažíme přiblížit.

3.1 Variabilita procesu

Abychom mohli použít metriky pro predikci nebo jako základ ke zkvalitňování procesů, musíme nejdříve zajistit stabilitu procesu. To je situace, kdy se všechna získaná data pohybují v předem stanovených mezích a po delší dobu zůstávají relativně neměnná. V opačném případě je rozptyl naměřených hodnot natolik velký, že nelze proces kontrolovat a je nutné podniknout kroky k jeho zkvalitnění. Hlavní příčinou nestability procesu je jeho variability, kterou popisuje následující podkapitola.

K tomu, abychom mohli zvyšovat kvalitu procesu nebo umět předpovídat jeho chování, musíme zajistit jeho stabilitu. Stabilita je podmíněna přijatelnými odchylkami v naměřených datech, která z pochopitelných důvodů vykazují během času různé nuance. Tyto odchylky mají dvě zásadní příčiny:

- jsou způsobeny samotnou podstatou procesu
- jsou způsobeny odchylkami, které mají přesnou příčinu a lze jim předcházet



Obrázek 3-1 Příčiny variability procesu

Odchylky, které mají konkrétní příčinu, jsou nejčastěji dány proměnlivostí složek, které tvoří proces. Například kvalita vstupních surovin má zásadní vliv na kvalitu celého výrobku. Proto jakákoliv odchylka na vstupu se projevuje také na výstupu. Tyto příčiny jsou pochopitelně důležité při zvyšování kvality a stability procesu.

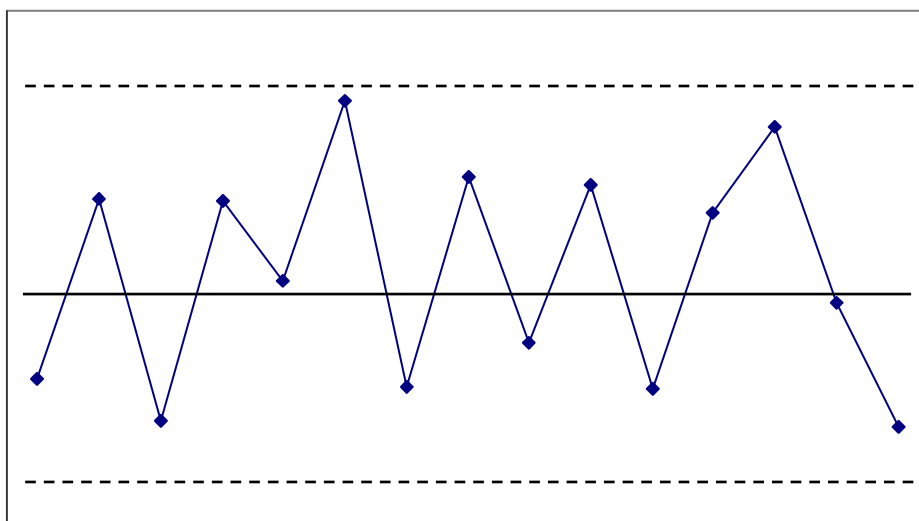
Co se týká odchylek způsobených samotnou podstatou procesu, ty nelze odstranit a budou se vyskytovat v naměřených datech vždy. Jejich význam a celkový dopad je však poměrně menší. Je velmi obtížné popsat a vysvětlit, jak vlastně vznikají. K tomuto účelu uvedeme příklad. Předpokládejme proces, který je konstantní. Znamená to, že jsme odstranili všechny chyby a odchylky, které mají konkrétní příčinu. Zajistili jsme například, že všechny vstupy procesu budou při každém jeho provedení vždy naprosto stejné, na procesu se budou podílet stejní lidé a budou používat stejné nástroje. I přesto bude výsledek procesu pokaždé trochu jiný. Tato odlišnost plyne právě ze samotné podstaty procesu. Hlavní význam v ní má lidský faktor, kdy člověk nedokáže vykonat tutéž činnost dvakrát naprosto stejně, což je příčinou změn. Jsme prostě limitováni tím, čeho chceme

dosáhnout. I zdánlivě jednoduché procesy budou vždy z jejich samotné podstaty vykazovat odlišné výsledky a s tímto faktem se musíme smířit.

Celková míra variability je tedy dána součtem odchylek plynoucích ze samotné podstaty procesu a odchylek, které mají konkrétní příčinu a ty se budeme snažit odstranit.

3.1.1 Vyhodnocování stability procesu

Význam řízení procesů spočívá v zajištění toho, aby všechna data byla v rámci stanovených limitů (viz. graf 3.1). Způsobům výpočtu těchto mezí bude věnována samostatná kapitola. Zatím předpokládejme, že umíme tzv. kontrolní meze určit. Pokud nová data tyto hranice překročí, dostaneme signál o nestabilitě procesu a o příležitosti podniknout kroky k nápravě.



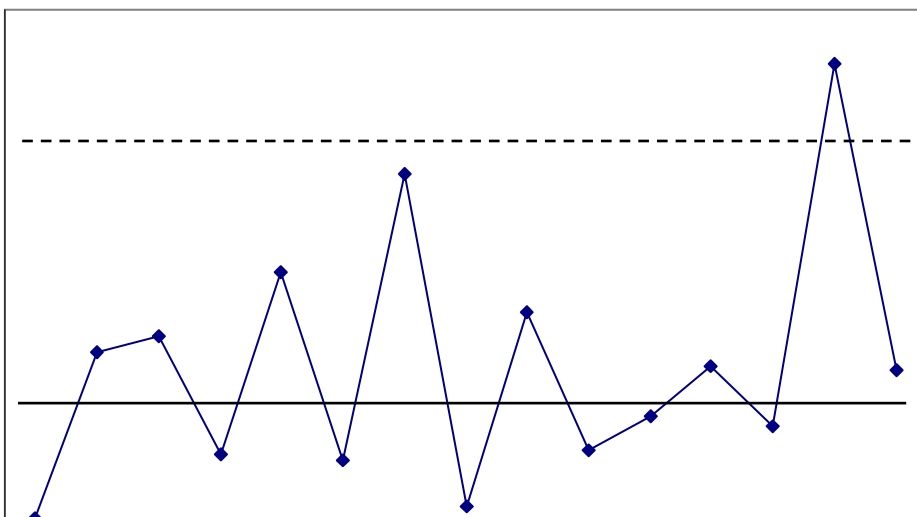
Graf 3.1 Graf zachycuje stabilní proces (naměřené hodnoty jsou v rámci limitů).

Pro vyhodnocování stability je nutná dobrá znalost procesů a definic všech metrik, které se pojí s daným procesem. Zároveň je také zapotřebí mít představu o tom, jak se má daný proces chovat nebo ještě lépe, mít historická data vykazující stabilní výkonnost, abychom mohli porovnávat. Vyhodnocování stability také v sobě zahrnuje proces identifikace obou typů odchylek (viz. kapitola 3.1 Variabilita procesu) a reakci na ně.

Základem statistické analýzy chování procesu jsou kontrolní grafy. Ty obsahují středovou linii, která nejčastěji odpovídá průměru naměřených hodnot, a kontrolní limity na obou stranách středové linie. Ty jsou vypočítány na základě naměřených dat. Konkrétní metody výpočtu hodnot těchto mezí budou popsány v následujících kapitolách. Pouze poznamenejme, že někdy nemusí kontrolní grafy obsahovat například spodní hranici limitu. Takováto situace nastane, když měříme počet chyb nebo počet zmetků. V takovémto okamžiku nepotřebujeme logicky spodní mez, protože tou je pochopitelně nula nebo-li žádný výskyt chyb (viz. Graf 3.2 Graf zachycuje příklad použití pouze jedné hraniční linie.).

Data, zobrazovaná těmito grafy, představují hodnoty jednotlivých měření v čase. Pokud některá z nich překročí horní nebo dolní mez, dostáváme okamžitě signál o nestabilitě procesu.

Hodnoty středových a kontrolních linií nelze vybrat libovolně, jelikož jejich účel je odhalit skutečné chování procesu. Kdyby totiž byly stanoveny ručně, reprezentovaly by spíš požadavek na to, jak se má daný proces chovat.



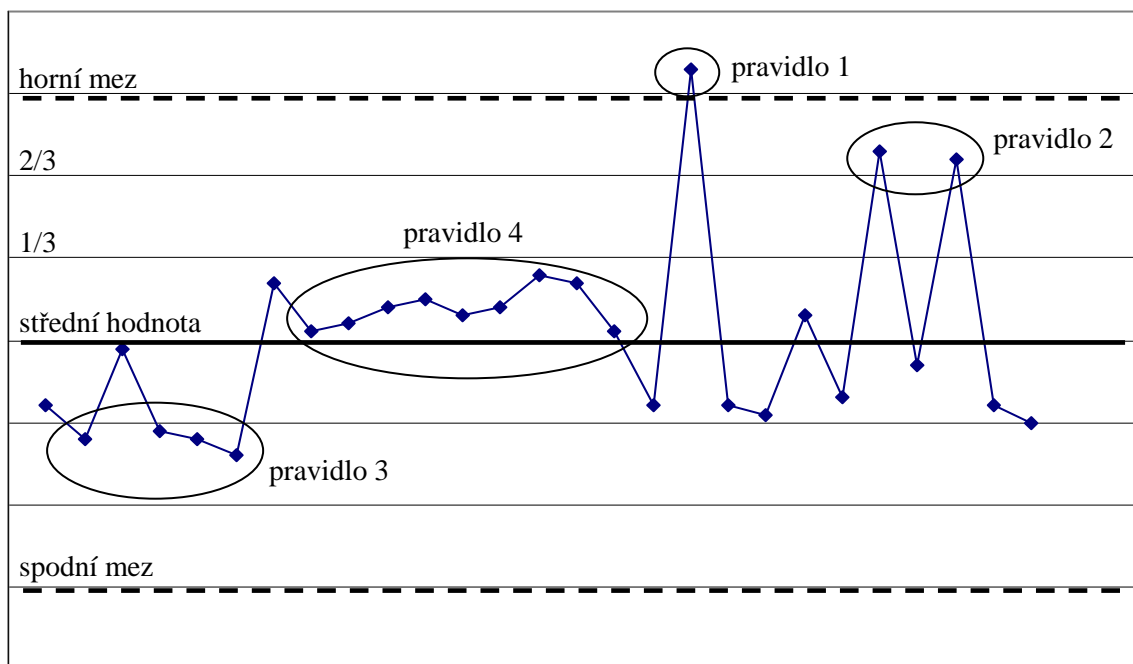
Graf 3.2 Graf zachycuje příklad použití pouze jedné hraniční linie.

Přesné techniky, pomocí kterých lze stanovit limity a středové hodnoty, budou blíže popsány v následujících kapitolách.

Nyní uvedeme základní čtyři pravidla detekce nestability procesu [WEE58, WHE92, WHE95] (viz. Obrázek 3-2 Příklad použití čtyř testů nestability).

1. jeden bod překračuje limity
2. alespoň dva body ze tří po sobě jdoucích překračují hranici 2/3 limitu
3. alespoň čtyři body z pěti po sobě jdoucích překračují hranici 1/3 limitu
4. alespoň osm po sobě jdoucích hodnot leží na stejné straně středové linie

Proces vykazuje nestabilitu v okamžiku, kdy se naplní i jediná z podmínek, díky čemuž můžeme získávat daleko přesnější představu o jeho chování. Na druhou stranu však nasazení všech čtyř testů s sebou nese i jisté komplikace. Především se zvýší počet falešných poplachů. Podle studií [CHA87] se může průměrný počet měření mezi dvěma falešnými poplasy zkrátit z 370 až na 91,25, což představuje čtyřnásobný nárůst počtu falešných poplachů.



Obrázek 3-2 Příklad použití čtyř testů nestability (převzato z [FLO99])

V praxi pochopitelně existuje daleko více podobných testů. Mají však všechny jednu společnou věc. Jejich nasazení do procesu analýzy se musí předem zvážit. Totiž v okamžiku, kdy jsou data zobrazená například v grafu, je na takové úvahy pozdě. Výsledky mohou být ovlivněny a celkově se také může zvýšit počet falešných poplachů. Člověk má totiž tendenci posuzovat věci jinak, pokud zná výsledek.

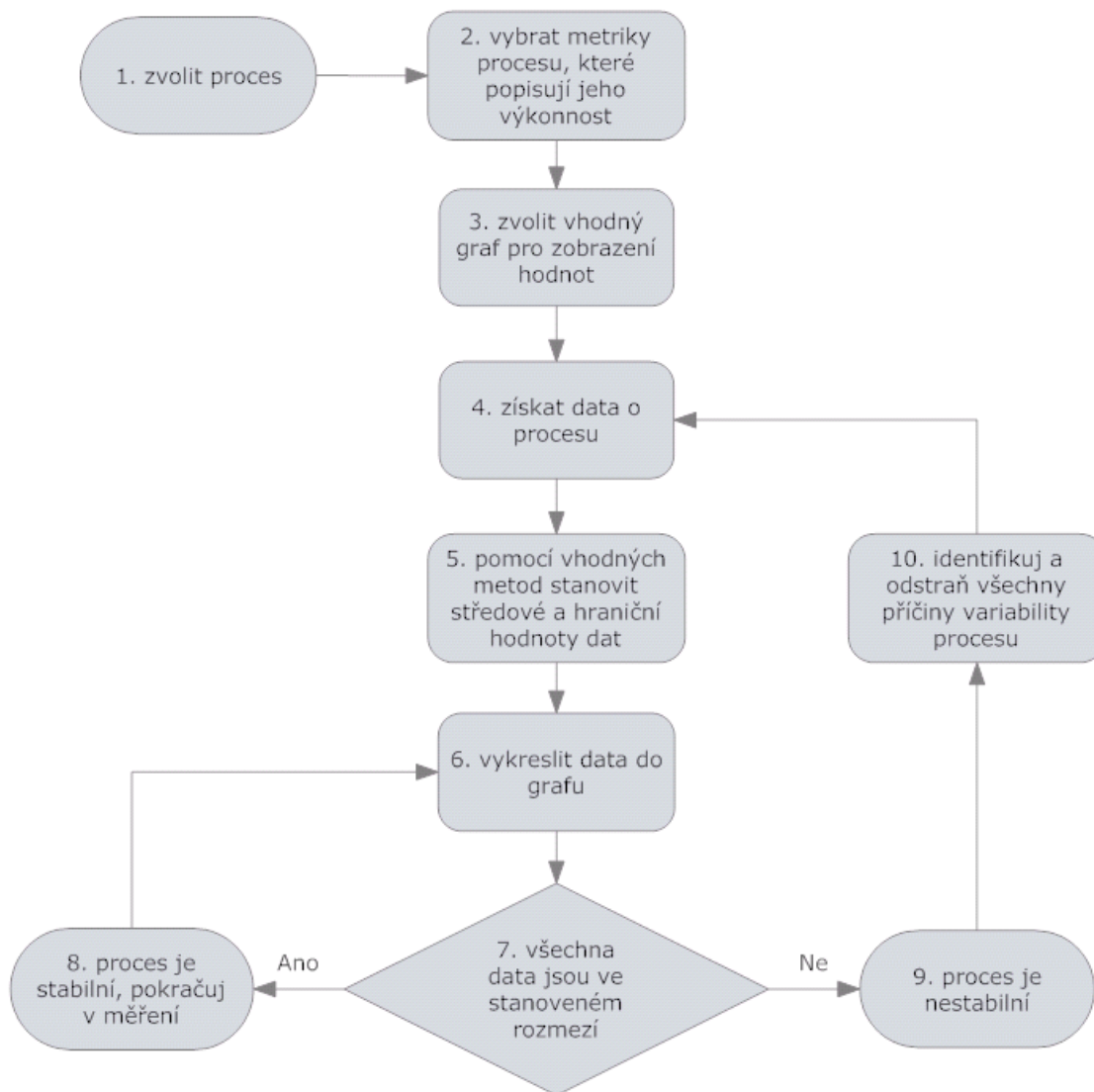
Nyní se zaměříme na samotný proces vyhodnocování stability. Ten využívá metrik procesů nebo metrik produktů ke stanovení stability procesu. Obsahuje celkem deset kroků, které zachycuje Obrázek 3-3 Kroky vyhodnocování stability procesu (převzato z [FLO99]).

Prvním krokem je výběr procesu, který chceme analyzovat z hlediska jeho stability.

Ve druhém kroku vybereme takové metriky, které nejvíce odpovídají cílům, kterých chceme dosáhnout. Mohou to být metriky procesu nebo jeho produktu.

Ve třetím kroku zvolíme vhodný graf, pomocí kterého budeme později získaná data analyzovat. Tento krok je nutné provést ještě před zahájením samotného měření, jelikož má vliv na následující měření.

Jako čtvrtý krok následuje měření procesu. Doba, po kterou data získáváme, je závislá na definici metriky, ve které je uvedena frekvence s jakou jsou data pořizována a také na typu vybraného grafu.



Obrázek 3-3 Kroky vyhodnocování stability procesu (převzato z [FLO99])

Pátým krokem je stanovení hraničních a středových hodnot dat pomocí vhodných metod. Jejich výběr závisí především na typu použitého grafu a typu dat.

V šestém kroku jsou data získaná v kroku čtyři zobrazena pomocí grafu.

V sedmém kroku dojde k porovnání dat. Zkoumá se, jestli jsou všechna data v rámci stanovených limitů. Pokud ano, pokračuje se krokem osm. Pokud tomu tak není a data splňují například některý z výše uvedených testů nestability procesu, pak se pokračuje krokem devět.

Krok osm znamená, že všechna zobrazená data splňují podmínku stability procesu. Dále se pokračuje v měření a opětovném zobrazování získaných hodnot do grafu v kroku šest.

Devátý krok znamená, že data jsou příliš variabilní a proces není tudíž stabilní. Důvod těchto odchylek musí být odhalen.

V desátém kroku jsou hledány příčiny nestability procesu. Vychází se například z definic metrik a jejich vazbě na konkrétní atributy procesu nebo jeho produktu, zkoumají se vzájemné vztahy a provádí se další pozorování za účelem přesného určení příčin nestability procesu. Pokud jsou v tomto kroku identifikovány odchylky procesu, které mají konkrétní příčinu, nejsou tedy způsobené samotnou podstatou procesu (viz. kapitola 3.1 Variabilita procesu), je nutné je odstranit a provést nové měření – krok čtyři.

V okamžiku, kdy jsme přesvědčeni o tom, že se daný proces chová stabilně, můžeme použít limity vypočítané v bodě pět k predikci jeho budoucího chování. Naopak, pokud proces vykazuje známky nestability, jsou takto stanovené limity pro předvídání chování procesu nepoužitelné. Je nutné v takovémto případě najít příčiny nestability, odstranit je a limity znovu vypočítat.

3.2 Grafy chování procesů

Tato kapitola přináší přehled všech typů grafů, které lze použít při analýze chování procesů. Zároveň všechny grafy rozděluje do dvou kategorií a to podle toho s jakými daty pracují na grafy dat proměnných a grafy kategorických data. Význam těchto pojmů je popsán v následující podkapitole. Je zde také podrobněji popsán význam a použití metody logického seskupování dat (Rational subgrouping [FLO99]).

3.2.1 Základní typy dat

Před samotným výkladem je nutné zavést rozdělení dat na dva základní typy [ZEN06]:

- data proměnných
- kategorická data

Data proměnných (někdy také nazývaná data měření) jsou obvykle spojena s měřením spojitých jevů. Příkladem může být délka, čas, hmotnost, vynaložené úsilí, zatížení procesoru, cena a pochopitelně mnohé další.

Oproti tomu kategorická data jsou velmi často spojována s počtem výskytů nebo s tím, zda nějaká veličina splňuje jisté kritérium nebo nabývá hodnoty z množiny daných atributů. Příkladem těchto dat může být počet nalezených vad, počet řádků jistého typu příkazu ve zdrojovém kódu, počet řádků komentářů a další.

Umět správně rozlišit o který typ dat se jedná je nutné již ve fázi definování metrik. Na základě naší volby pak používáme odlišné typy grafů, které pracují s naměřenými daty. Pro úplnost ještě ilustrujeme způsob rozdělení dat na jednoduchém příkladě. Je nutné si uvědomit, že ne vždy jsou data proměnných spojena se spojitými jevy. Budeme uvažovat počet dnů potřebných ke splnění jistého úkolu. Mohlo by se zdát, že tato data jsou kategorická, jelikož se jedná o počty výskytů jisté události a zároveň se nejedná o spojitou veličinu. Opak je pravdou. V této situaci je celkový počet dnů

prakticky neomezený a naměřená data tak mohou nabývat libovolných hodnot. Jiná situace je, když budeme uvažovat o počtu dnů v měsíci, například pro výpočet procent pracovních dnů. Zde mohou naměřená data nabývat jen omezených hodnot a z tohoto důvodu se na ně budeme dívat jako na data kategorická.

Pokud tedy shrneme výše uvedené. Dělení data se řídí částečně podle jejich definice, ale hlavním vodítkem zde vždy musí být způsob jejich sběru a použití. Pokud budeme měřit počet výskytů jistého jevu, který může nabývat neomezených hodnot, pak o těchto datech budeme uvažovat jako o datech proměnných a podle tohoto také budeme přizpůsobovat výběr grafů vhodných pro analýzu. Pochopení rozdílů mezi typy dat je zásadní pro další výklad.

3.2.2 Logické seskupování dat

Tato podkapitola objasňuje základní princip, který se uplatňuje při práci takřka se všemi typy grafů. Jedná se o logické seskupování dat, které může zásadním způsobem negativně ovlivnit výsledek analýzy chování procesu pokud mu není věnovaná dostatečná pozornost.

Logické seskupování dat znamená transformaci (sloučení) naměřených dat do skupin o velikosti n . Lze si představit například seskupení všech měření z daného týdne do skupin o velikosti 5. Každá takto vytvořená skupina musí obsahovat homogenní data, to znamená taková data, ve kterých se vykytuje variabilita způsobená pouze samotnou podstatou procesu. Toto je první požadavek logického seskupování. V našem příkladě to znamená, že data měřená od pondělí do pátku byla měřená za stejných podmínek a rozdíly se projeví pouze mezi jednotlivými týdny.

Seskupením dat tímto způsobem také říkáme, že nás nezajímá variabilita procesu měnící se ze dne na den, ale z týdne na týden. A právě změny chování procesu mezi jednotlivými týdny jsou tím, co chceme analyzovat, což splňuje druhou podmínku logického seskupování. Ta říká, že data musejí být vždy seskupována takovým způsobem, aby výsledek analýzy co nejlépe korespondoval s hlavními cíli zvyšování kvality procesů.

Logické seskupování budeme dále používat téměř u všech typů grafů. Jeho dva základní principy jsou jednou z hlavních myšlenek analýzy dat založené na statistickém přístupu.

3.3 Grafy dat proměnných

Tato podkapitola přináší přehled grafů používaných pro analýzu chování procesů s využitím dat proměnných. Pro úplnost je tento typ dat spojen s měřením spojitých jevů nebo při měření počtu výskytů jistého jevu, který může nabývat neomezených hodnot.

3.3.1 X-bar/R graf

První dvojice grafů, kterými jsou X-bar/R graf, se používá v situaci, kdy provádíme několik měření v poměrně krátkém čase a za stejných podmínek a toto měření několikrát opakujeme. Pokud jsou totiž data sbírána za stejných podmínek, tvoří skupinu (viz. Logické seskupování dat), která obsahuje pouze odchylky způsobené samotnou podstatou procesu a tudíž splňují první podmínku logického seskupování.

První z grafů (X-bar) odpovídá na otázku, do jaké míry se mohou lišit průměrné hodnoty jednotlivých skupin s ohledem na rozdíly uvnitř těchto skupin. Druhý graf (R graf) naproti tomu zobrazuje míru variability uvnitř jednotlivých skupin a odpovídá tak na otázku, zda jsou rozdíly uvnitř skupin konstantní v čase.

Jejich bližší význam si ukážeme na postupu výpočtu kontrolních mezí (HM – horní mez, DM – dolní mez, SH – střední hodnota). Mějme k měření, z nichž každé tvoří skupina dat o velikosti n . V prvním kroku vypočítáme průměr \bar{X}_k a rozsah R_k každé z k skupin.

$$\bar{X}_k = \frac{X_1 + X_2 + \dots + X_n}{n} \quad R_k = |X_{MAX} - X_{MIN}| \quad (1)$$

Dále vypočítáme celkový průměr dat $\bar{\bar{X}}$ a průměrný rozsah skupin \bar{R} jako:

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_k}{k} \quad \bar{R} = \frac{R_1 + R_2 + \dots + R_k}{k} \quad (2)$$

Na základě těchto hodnot můžeme přistoupit k výpočtu kontrolních mezí. Pro X-bar graf se hodnoty spočítají podle rovnice (3).

$$\begin{aligned} HM_{\bar{X}} &= \bar{\bar{X}} + A_2 \bar{R} \\ SH_{\bar{X}} &= \bar{\bar{X}} \\ DM_{\bar{X}} &= \bar{\bar{X}} - A_2 \bar{R} \end{aligned} \quad (3)$$

Pro R graf se hodnoty kontrolních mezí počítají podle rovnice (4).

$$\begin{aligned} HM_{\bar{R}} &= D_4 \bar{R} \\ SH_{\bar{R}} &= \bar{R} \\ DM_{\bar{R}} &= D_3 \bar{R} \end{aligned} \quad (4)$$

Hodnoty konstant A_2, D_3, D_4 , které dosazujeme do rovnic, lze získat z následující tabulky.

n	d_2	A_2	D_3	D_4
2	1,128	1,880	-	3,268
3	1,693	1,023	-	2,574
4	2,059	0,729	-	2,282
5	2,326	0,577	-	2,114
6	2,534	0,483	-	2,004
7	2,704	0,719	0,076	1,924
8	2,847	0,373	0,136	1,864
9	2,970	0,337	0,184	1,816
10	3,078	0,308	0,223	1,777

Tabulka 3-1: Konstanty pro výpočet kontrolních mezí X-bar a R grafu (převzato z [WHE92])

Hodnoty ve sloupci d_2 jsou vypočítány na základě matematických vzorců na základě statistických metod [HAR60]. Z nich se pak následně pomocí rovnic vypočítávají ostatní koeficienty v tabulce [FLO99].

Pro analýzu chování procesu pak hodnoty kontrolních linií zaneseme do grafu a použijeme testy stability uvedené výše (viz. Obrázek 3-2). Z výše uvedených rovnic je patrné, že X-bar graf zobrazuje průměrné hodnoty jednotlivých skupin měření. Vychýlení dat v některé ze skupin se proto vždy zobrazí také v grafu. Příkladem těchto situací mohou být různé výpadky systémů nebo naopak zvýšený počet požadavků, které není možné uspokojit, v důsledku čehož se může například prodloužit průměrná doba jejich zpracování.

Při analýze chování procesu však vždy používáme X-bar graf společně s R grafem. Jakákoliv známka nestability v R grafu nasvědčuje často o tom, že daný proces je ovlivňován nějakým skrytým podprocesem. Pokud však R graf odhalí nestabilní chování, X-bar graf již dále neinterpretujeme.

Příkladem použití této dvojice grafů může být analýza počtu chyb. Řekněme, že budeme mít šest modulů jisté aplikace a před vydáním její nové verze budeme zkoumat počet chyb v těchto modulech. Využijeme seskupování, kdy každá skupina bude mít velikost šest (počet modulů) a takto naměříme počet chyb pro více verzí aplikace. Takto upravená data splňují obě podmínky logického seskupování – jsou homogenní a odpovídají cílům analýzy.

3.3.2 X-bar/S graf

Předchozí R graf lze použít pouze v případě, kdy velikost skupin je menší než 10. Pro větší počet měření v dané skupině použijeme místo něj S graf, který poskytuje ve výsledku užší kontrolní meze a je tak citlivější při odhalování skrytých příčin nestability procesu. Jeho nevýhodou může být náročnější způsob výpočtu mezí, který používá mocniny.

Prvním krokem je výpočet standardní odchylky S_k z k skupin dat, každé o velikosti n . Poté se vypočítá průměrná odchylka dat \bar{S} .

$$S_k = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad \bar{S} = \frac{\sum_{i=1}^k S_i}{k} \quad (5)$$

Na základě těchto hodnot můžeme přistoupit k výpočtu kontrolních mezí. Pro X-bar graf se hodnoty spočítají podle rovnice (6).

$$\begin{aligned} HM_{\bar{X}} &= \bar{\bar{X}} + A_3 \bar{S} \\ SH_{\bar{X}} &= \bar{\bar{X}} \\ DM_{\bar{X}} &= \bar{\bar{X}} - A_3 \bar{S} \end{aligned} \quad (6)$$

Pro S graf se hodnoty kontrolních mezí počítají podle rovnice (7)

$$\begin{aligned} HM_{\bar{S}} &= B_4 \bar{S} \\ SH_{\bar{S}} &= \bar{S} \\ DM_{\bar{S}} &= B_3 \bar{S} \end{aligned} \quad (7)$$

Následující tabulka obsahuje hodnoty konstant pro jednotlivé velikosti skupin.

n	A_3	B_3	B_4
6	1,287	0,030	1,970
7	1,182	0,118	1,882
8	1,099	0,185	1,815
9	1,032	0,239	1,761
10	0,975	0,284	1,716
11	0,927	0,322	1,678
12	0,886	0,354	1,646
13	0,850	0,382	1,619
14	0,817	0,407	1,593
15	0,789	0,428	1,572

Tabulka 3-2: Konstanty pro výpočet kontrolních mezí S grafu (převzato z [WHE92])

Výpočet konstant je opět položen na základech matematiky a statistiky [HAR60] [FLO99].

S graf lze použít i pro skupiny dat větší než 15. Existují vzorce pro výpočet příslušných faktorů [WHE92], je však nutné počítat s tím, že s rostoucí velikostí skupin vzrůstá také senzitivita grafu, jelikož kontrolní linie se zužují.

3.3.3 X-bar/mR graf

Pro zobrazení individuálních hodnot měření používáme X-bar/mR graf, který se také někdy nazývá graf samostatných hodnot, zkráceně XmR. Ne vždy totiž může být praktické seskupovat měření. Může se tím zakrýt samotná podstata měření nebo v horším případě zakrýt nežádoucí chování procesu. Uplatnění proto tento graf nalezne v situaci, kdy jsou jednotlivá měření používána samostatně pro hodnocení chování procesu. Může to být v situaci, kdy jsou měření nepravidelná nebo jsou mezi nimi větší časové rozestupy a podmínky při měření se mění, což odporuje prvnímu principu logického seskupování. Velikost skupin je v takovém případě $n = 1$.

Pomocí grafu lze zobrazovat jak spojité, tak i diskrétní veličiny. Pro druhou variantu však musíme zajistit, aby celkový počet možných hodnot byl minimálně 5 [FLO99]. V takovém případě můžeme o diskrétních datech uvažovat jako o spojitých. Tento graf můžeme použít například při analýze počtu hodin strávených na odstraňování chyb (spojitá veličina) nebo při analýze počtu požadavků zákazníka (diskrétní data).

Pro výpočet kontrolních mezí je použita naprosto jiná logika. Základní myšlenkou je sloučení vždy dvou po sobě jdoucích měření do skupin o velikosti $n = 2$. Absolutní hodnota rozdílu těchto dvou měření mR_i se použije pro výpočet celkové odchylky procesu \overline{mR} (moving range) a to podle rovnice (8).

$$mR_i = |X_{i+1} - X_i|, 1 \leq i \leq r, r = k - 1$$

$$\overline{mR} = \frac{\sum_{i=1}^{i=r} mR_i}{r} \quad (8)$$

Pro výpočet kontrolních mezí použijeme Tabulka 3-1, ze které vybereme konstantu $d_2 = 1,128$ (korekční faktor) pro $n = 2$ a výpočet pro X-bar provedeme podle rovnice (9).

$$HM_{\bar{X}} = \bar{X} + \frac{3\overline{mR}}{d_2} = \bar{X} + 2,660\overline{mR}$$

$$SH_{\bar{X}} = \bar{X} = \frac{1}{k} \sum_{i=1}^k X_i \quad (9)$$

$$DM_{\bar{X}} = \bar{X} - \frac{3\overline{mR}}{d_2} = \bar{X} - 2,660\overline{mR}$$

Kontrolní meze mR grafu vypočítáme podle rovnice (10), do které dosadíme hodnoty z Tabulka 3-1. Je nutné poznamenat, že dolní mez mR grafu je rovna nule, jelikož v dané tabulce není definována hodnota koeficientu D_3 pro $n = 2$.

$$HM_{mR} = D_4 \overline{mR} = 3,268\overline{mR}$$

$$SH_{mR} = \overline{mR} \quad (10)$$

$$DM_{mR} = D_3 \overline{mR} = 0$$

Fakt, že pro výpočet kontrolních mezí X-bar/mR grafu se používá průměr naměřených hodnot, může mít vliv na výsledky analýz. Odlehle hodnoty v datech mohou rozšířit kontrolní meze (posunout dále od sebe) a tím snížit citlivost grafu k detekci nestability. Pro tento případ existuje ještě jedna varianta výpočtu hodnoty \overline{mR} , která je označována jako \overline{mmR} (median moving range) a místo průměru hodnot uvažuje medián. Odpovídající graf je pak označován jako X-bar/mmR graf.

V seřazeném souboru dat hodnotu \overline{mmR} vypočítáme tak, že pokud má daný soubor lichý počet záznamů, pak je hodnota \overline{mmR} rovna hodnotě prostředního prvku a pokud je počet prvků sudý, pak je hodnota \overline{mmR} rovna průměru prostředních dvou. Postup výpočtu kontrolních mezí X-bar/mmR grafu je uveden v rovnicích (11) a (12). Hodnota faktoru $d_4 = 0,954$ (převzato z [FLO99]).

$$HM_{\bar{X}} = \bar{X} + \frac{3\overline{mmR}}{d_4} = \bar{X} + 3,145\overline{mmR}$$

$$SH_{\bar{X}} = \bar{X} = \frac{1}{k} \sum_{i=1}^k X_i \quad (11)$$

$$DM_{\bar{X}} = \bar{X} - \frac{3\overline{mmR}}{d_4} = \bar{X} - 3,145\overline{mmR}$$

$$HM_{\overline{mmR}} = D_6 \overline{mmR} = 3,865\overline{mmR}$$

$$SH_{\overline{mmR}} = \overline{mmR} \quad (12)$$

Tento typ grafu není tak citlivý na odhalování nestabilního chování procesu jako X-bar/R graf. Výsledky, které poskytuje, jsou však velmi podobné. Výborně se tento typ grafu hodí pro identifikaci skrytých vzorů chování, které se projeví daleko dříve než v jiných typech grafů. Této problematice bude věnována samostatná kapitola (viz. kapitola 3.6.1 Sledování trendů). Tento graf se také často používá pro posuzování míry vspělosti procesu. Při analýze chování procesu za použití mR (resp. mmR) grafu se nedoporučuje používat čtyři testy nestability, které byl popsán dříve [WHE92].

3.3.4 MAMR graf

Posledním typem grafu, který lze použít pro analýzu dat proměnných je MAMR graf (moving average and moving range). Využití nalezne především při hlubší analýze trendů daného procesu. Všechny předešlé grafy se zaměřovaly na vzájemný vztah jednotlivých měření a porovnávaly je mezi sebou. MAMR graf naproti tomu lze použít například při analýze výkonnosti procesu, kdy manažerům poskytuje cenná data, které napovídají o tom, jak moc daný proces drží svoji vzestupnou (nebo naopak sestupnou) tendenci výkonnosti. Tento graf proto může být použit při rozhodování a plánování projektu.

Postup výpočtu kontrolních mezí je následující. Mějme k individuálních měření. Zvolíme velikost skupiny $n \geq 2$. Mějme však na paměti, že hodnotu n musíme přizpůsobit velikosti vzorku dat k . Pro malé hodnoty k musí být hodnota n rozumně malá a naopak při velké hodnotě n se snižuje homogenita uvnitř skupin, což odporuje prvnímu principu logického seskupování.

Zapíšeme hodnoty měření do tabulky o n řádcích, vždy na následujícím řádku posunuté o jednu doprava. Následně vypočítáme průměr hodnot \overline{mX}_k (moving average) ve sloupcích, které obsahují všechny hodnoty. Tento postup je znázorněn v Tabulka 3-3. Pro tento příklad uvažujeme $n = 2, k = 4$ (hodnoty měření 24, 23, 15, 18)

	k1	k2	k3	k4	k5
n-1	24	23	15	18	
n		24	23	15	28
mX	-	23,5	19	16,5	-

Tabulka 3-3 Ukázka výpočtu MAMR grafu.

Výpočet \overline{mX}_k lze zapsat formálně následujícím způsobem:

$$\begin{aligned}\overline{mX}_k &= \frac{X_k + X_{k-1}}{2} && \text{pro } k \geq n = 2 \\ \overline{mX}_k &= \frac{X_k + X_{k-1} + \dots + X_{k-n+1}}{n} && \text{pro } k \geq n \geq 3\end{aligned}\quad (13)$$

Poté vypočítáme celkový průměr hodnot jako:

$$\overline{mX} = \sum_{k=n}^{k=N} \frac{\overline{mX}_k}{N - n + 1} \quad (14)$$

Výpočet kontrolních limitů MAMR grafu probíhá naprosto stejně jako u X-bar/R grafu, tzn. podle rovnic (3) a (4). Při analýze chování procesu za použití MAMR grafu se nedoporučuje používat čtyři testy nestability [WHE92].

3.4 Grafy kategoričkých dat

Tato podkapitola přináší přehled grafů používaných pro analýzu chování procesů s využitím kategoričkých dat. Přináší popis tří grafů: C, U a XmR grafu, které lze pro tento typ dat použít.

Pro úplnost připomeňme, že se jedná o data, která typicky nabývají diskretních hodnot a kardinalita oboru hodnot je malá. Grafy použijeme například v situaci, kdy budeme analyzovat počet chyb v modulu, počet neplánovaných výpadků serveru, počet chyb na jeden tisíc řádků kódu a další podobné případy.

V souvislosti s tímto typem dat je vždy nutné se zabývat problematikou normalizace. Kardinalita zkoumaného prostoru se může obecně mezi jednotlivými měřeními lišit. Pokud budeme například analyzovat počet chyb v modulu, výsledek bude vždy závislý na jeho velikosti. Stejně tak se zvyšujícím se počtem řádků zdrojového kódu se zvyšuje pravděpodobnost výskytu chyby, což pochopitelně znesnadňuje porovnávání takovýchto dat navzájem mezi sebou.

S otázkou normalizace také úzce souvisí problematika správné definice metrik, která byla popsána v předchozích kapitolách. Pokud budeme chtít analyzovat chování procesu pomocí metrik, jejichž obor se obecně liší, je nutné do definice takovéto metriky zahrnout také sběr dat, která budou popisovat právě velikost zkoumaného prostoru v době měření. Bez úplné definice by totiž nebylo možné provádět normalizaci. Příkladem může být počet chyb ve zdrojovém kódu, kdy ke každému měření připojujeme také informaci o tom, jaký byl celkový počet řádek kódu. Jedině tak můžeme data normalizovat a jednotlivá měření mezi sebou porovnávat. V našem případě bychom přepočítali počet chyb na jeden tisíc řádek zdrojového kódu.

Dále tedy budeme rozdělovat grafy do dvou skupin a to podle toho, zda je velikost zkoumaného prostoru konstantní nebo ne:

- C graf a XmR graf (konstantní velikost prostoru)
- U graf a XmR graf (proměnlivá velikost zkoumaného prostoru)

3.4.1 C graf

Tento typ grafu je použitelný pro analýzu chování procesu, který popisují kategoričká data a u všech měření je velikost zkoumaného prostoru konstantní. Příkladem může být vyhodnocování počtu chyb

v úsecích kódu stejné délky. Společně s U grafem se jedná o velmi účinnou metodu analýzy chování procesu, která je vhodná zejména u softwarových procesů.

Při použití tohoto grafu se však musíme ujistit, že data popisující chování procesu odpovídají Poissonovu rozdělení pravděpodobnosti. To zahrnuje především fakt, že jednotlivá měření jsou na sobě naprosto nezávislá, události mají diskrétní charakter a vyskytují se poměrně zřídka vzhledem k celkovému počtu pokusů [WHE92][FAJ]. Jinými slovy lze říct, že počet úspěšných pokusů je zanedbatelně malý oproti celkovému počtu pokusů [FLO99]. Pokud tedy budeme zkoumat počet chyb ve zdrojovém kódu, je možné považovat jednotlivé řádky za pokusy, ve kterých může dojít k výskytu události, v našem případě k výskytu chyby ať už syntaktické, sémantické nebo logické. Výskyt této události lze popsat právě Poissonovým rozdělením.

Označme celkový počet úspěšných událostí (např. nalezení chyby v kódu) jako s a celkový počet zkoumání (počet řádek kódu) jako e . Pro výpočet kontrolních mezí C grafu použijeme následující rovnice:

$$\bar{c} = \frac{s}{e} \quad \begin{aligned} HM_c &= \bar{c} + 3\sqrt{\bar{c}} \\ SH_c &= \bar{c} \\ DM_c &= \bar{c} - 3\sqrt{\bar{c}} \end{aligned} \quad (15)$$

Vždy je však nutné ujistit se, že při nasazení tohoto grafu jsou splněny podmínky Poissonova rozdělení tak, jak byly popsány výše. Pokud tento graf budeme chtít použít pro analýzu chování procesu, musíme mít vždy na paměti, že všechny výpočty s ním spojené jsou založené na teoretickém modelu a jeho možnosti jsou proto pouze omezené. Pokud data zobrazená v tomto grafu vykazují známky nestability, interpretujeme výsledky a hledáme příčiny tohoto chování stejným způsobem, jako u X-bar grafu.

3.4.2 U graf

Tento typ grafu je velmi podobný C grafu. Naměřená data musejí opět odpovídat Poissonovu rozdělení, rozdíl je však v tom, že kardinalita zkoumaného prostoru a_i je u každého měření jiná. Výpočet kontrolních mezí proto předchází normalizace těchto naměřených dat. Výsledek každého měření c_i je dělen právě velikostí prostoru a_i :

$$u_i = \frac{c_i}{a_i} \quad (16)$$

Takto vypočítané hodnoty mohou být zaneseny do grafu. Pro výpočet kontrolních mezí však musíme nejdříve vypočítat celkovou průměrnou hodnotu jednotlivých měření \bar{u} a kontrolní meze vypočítat podle následujících rovnic:

$$\bar{u} = \frac{\sum c_i}{\sum a_i} \quad \begin{aligned} HM_u &= \bar{u} + 3\sqrt{\frac{\bar{u}}{a_i}} \\ SH_u &= \bar{u} \\ DM_u &= \bar{u} - 3\sqrt{\frac{\bar{u}}{a_i}} \end{aligned} \quad (17)$$

Je nutné poznamenat, že pro tento typ grafu nemají horní a dolní kontrolní mez konstantní hodnotu pro všechna měření. Z rovnice (17) je patrné, že kontrolní meze jsou vypočítávány pro každé měření zvlášť a mají tudíž různé hodnoty. Při používání tohoto grafu pro analýzu procesu platí stejné zásady, které byly zmíněny u C grafu.

3.4.3 XmR graf

XmR graf byl popisován již dříve v souvislosti s typem proměnných dat (viz. kapitola 3.3.3). Lze jej však použít i pro analýzu procesu, který popisují kardinální data. Jeho výhoda spočívá především v tom, že potřebuje daleko méně dat k výpočtu kontrolních mezí. Použijeme jej především v situaci, kdy si nejsme jisti přesným typem dat nebo potřebujeme mít výsledky analýzy co nejdříve. Grafy určené speciálně pro kategorická data (C a U graf) však poskytují daleko přesnější výsledky a jejich použití zajistí užší kontrolní linie a vyšší citlivost na odhalování nestabilit procesu.

Použitím tohoto grafu však zároveň zastíráme Poissonův model pravděpodobnosti a předpokládáme, že standardní odchylka dat je u všech měření konstantní. Z tohoto důvodu není zcela správné nasazovat tento graf pro analýzu výskytu chyb v kódu, jelikož tento případ zcela jasně vede na použití Poissonova modelu. Zároveň není vhodné tento graf používat v situacích, kdy se jednotlivé události vyskytují jen zřídka a nebo když je výsledek většiny měření nulový.

3.5 Technická stránka analýzy dat

Tato kapitola přináší podrobný přehled metod a technik, které se uplatňují při analýze dat. Diskutována je zde především otázka minimálního množství dat, které je dostačující pro zahájení analýz, dále je na příkladech popsána metoda postupné eliminace dat. Mezi pokročilejší techniky pak patří identifikace trendů v grafu, které je věnována samostatná podkapitola. Dále jsou zde diskutovány další doporučení, kterých je nutné se nejen při analýze dat držet. Závěr této kapitoly je věnován stručnému přehledu technik, které se uplatňují při zvyšování vspělosti procesu (*process capability*). Cílem této kapitoly je přinést teoretické i praktické poznatky z oblasti technických možností analýzy chování procesu pomocí grafů.

3.5.1 Hromadění dat

Při analýze chování procesu využíváme data, která nám poskytují metriky. Na začátku měření nemáme o procesu žádné informace. Až s přibývajícím časem se začnou data hromadit a je vhodný čas je analyzovat. Otázkou zůstává, jaké množství dat je dostačující k dosažení spolehlivých výsledků analýz.

Z předchozích kapitol vyplynulo, že množství dat je závislé na počtu měření k a na velikosti skupiny n . Pokud budeme chtít vypočítat kontrolní limity procesu, budeme potřebovat nejméně 3 až 4 měření ($k > 2$) [FLO99]. V takovémto případě můžeme proces analyzovat a hledat příčiny jeho nestability již krátce po započetí měření. Na velikosti skupiny n v takovémto případě nezáleží. Takto vypočítané kontrolní limity jsou pouze orientační a slouží především k tomu, aby bylo možné odhalit příčiny nestability procesu hned na začátku. V opačném případě bychom totiž museli provést přibližně 25 měření, abychom získali stabilní a spolehlivé meze. Tato technika brzké analýzy je velmi účinná a informace, které poskytuje jsou velmi cenné. Pomáhá hledat a odstraňovat příčiny nestability procesu hned na jeho začátku, což celkově urychluje jeho stabilizaci a je to první krok při zlepšování procesů.

Dalším krokem je zvyšování vspělosti procesu (*capability*). Jedná se například o situaci, kdy máme již dlouho fungující proces a k němu data z několika desítek měření. V takovém případě, budeme chtít vypočítané kontrolní meze použít k tomu, abychom určili míru vspělosti procesu, porovnali chování procesu vůči předem daným požadavkům nebo abychom byli schopni předikovat jeho chování v budoucnu. Více o zvyšování vspělosti procesu v následující kapitole.

Ve všech třech případech musíme mít ve vypočítané kontrolní meze velkou důvěru. Z tohoto důvodu také potřebujeme mít větší množství dat. V případě X-bar/R grafu je to minimálně 25 měření ($k > 24$) a pro výpočet spolehlivých kontrolních mezí u XmR grafu budeme potřebovat alespoň 40 měření ($k > 39$) [FLO99]. Důvod tohoto je zřejmý. Větší množství dat minimalizuje pravděpodobnost odchýlení kontrolních mezí od správných hodnot v situaci, kdy některá data budou vykazovat známky nestability a hodnoty těchto měření budou velmi odlehlé od průměru.

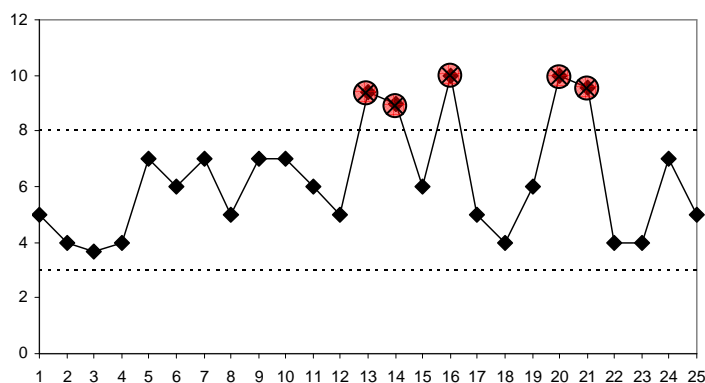
Dosud byla řeč pouze o tom, jaký vliv má počet měření k na vypočítané kontrolní meze. Nyní se zaměříme na velikost skupiny n . Volba počtu vzorků dat ve skupině má zásadní dopad na analýzu z pohledu citlivosti kontrolních mezí. Na jednu stranu je vhodné mít co největší skupiny dat, jelikož takto vypočítané kontrolní meze jsou užší a lze pomocí nich snáze odhalit nestabilitu procesu. Na druhou stranu se zvyšujícím se počtem vzorků dat ve skupině roste také riziko toho, že skupina nebude homogenní, což je první požadavek logického seskupování (viz. kapitola 3.2.2 Logické seskupování dat). Do dat se může dostat šum a odlehlé hodnoty, což je nepřijatelné. Při volbě velikosti n se snažíme, aby skupina byla co největší, ale zároveň, aby se v ní nikdy nevyskytly nehomogenní data, která by mohla ovlivnit výsledky analýz. Je daleko důležitější zachovat homogenitu skupin, než neúměrně zvyšovat velikost skupiny. Pokud se však stane, že v dané skupině dat jsou hodnoty s velkým rozsahem, je nutné mít na paměti, že i tato data mohou být v grafu signálem nestability procesu.

3.5.2 Metoda eliminace odlehlých dat

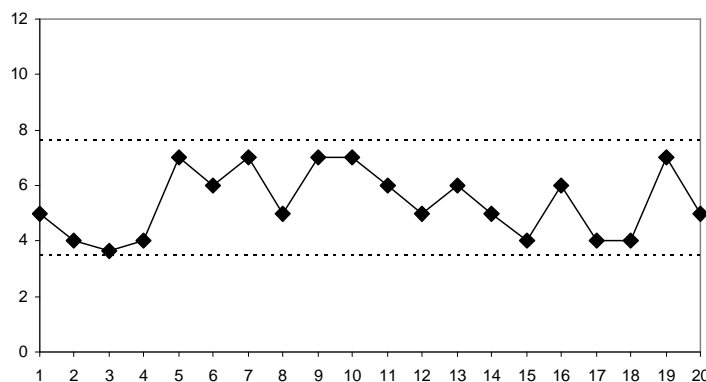
Tato podkapitola popisuje nejčastější postup při analýze chování procesu s využitím grafů, který je založený na postupné eliminaci odlehlých bodů a přepočítávání kontrolních mezí.

Hlavním cílem této metody je provést takové úpravy dat v grafu, abychom našli všechny příčiny nestability procesu, pokud možno i ty skryté. Základem je graf, který obsahuje všechna měření a na jejich základě vypočítané kontrolní meze. V jednotlivých krocích pak dochází k odstraňování odlehlých hodnot z grafu, tedy těch, které jsou mimo vypočítané kontrolní meze. Poté jsou ze zbývajících hodnot meze přepočítány. Nové meze jsou pochopitelně užší a citlivější na odhalování nestability procesu. Některé body se tak mohou dostat opět mimo ně. Provedeme jejich odstranění a znovu přepočítáme kontrolní meze. Tento postup můžeme opakovat až do té doby, dokud se všechny body v grafu nebudou nacházet v rámci kontrolních mezí.

Celý proces zachycuje Obrázek 3-4. Na něm je vidět XmR graf procesu s vynesnými daty a vypočítanými kontrolními limity. Postupnou eliminací ve dvou krocích vznikl graf na Obrázek 3-5, u kterého byly odstraněny všechny odlehlé hodnoty a ze zbývajících dat byly vypočítány kontrolní meze.



Obrázek 3-4 Graf chování procesu před eliminací bodů



Obrázek 3-5 Graf chování procesu po eliminaci odlehlých hodnot

Všechny eliminované body musíme následně podrobit analýze. Mohou totiž indikovat nestabilitu procesu, kterou se snažíme odstranit. Při používání metody postupné eliminace bodů není vhodné analyzovat každý jednotlivý bod zvlášť, ale naopak mezi nimi hledat závislosti a vzory tak, jak to je popsáno v následující kapitole.

S otázkou analýzy naměřených hodnot souvisí také problematika vypořádání se s nerepresentativními daty. To jsou takové hodnoty, které jsou pro proces naprosto netypické nebo mohou být jakkoliv nevhodné či špatně pořízené a jejich zahrnutí do analýzy by vedlo zcela jistě ke zkreslení výsledků. Pokud dokážeme taková data s jistotou identifikovat, můžeme je z analýzy vyloučit. Mohlo by se zdát, že takovýto zásah není správný, že takto můžeme přijít o data, která popisují netypické chování procesu a mohou se stát cenným zdrojem informací při analýze. Smyslem je však eliminovat pouze taková data, která s jistotou nepopisují chování procesu s ohledem na jeho podstatu.

3.6 Pokročilé techniky analýzy

Tato podkapitola přináší několik dalších rad a doporučení, která souvisejí s analýzou chování procesů. Zdůrazňuje především nutnost zaznamenávat dodatečná data při každém měření a diskutuje problematiku četnosti sbírání dat v souvislosti s volbou velikosti skupin. Věnuje se také analýze trendů v grafech a zmiňuje pojem vyspělost procesu.

K tomu, abychom mohli zvyšovat kvalitu procesů, musíme mít správná data a umět je dobře analyzovat. Hlavním cílem analýzy je nalezení příčin nestability procesu a jejich následné odstranění. Abychom byli schopni toto provést, musejí být výsledky jednotlivých měření správné a především by měly obsahovat další dodatečné informace, které by mohli být užitečné. Jinými slovy lze říct, že je důležité zaznamenávat společně s výsledky jednotlivých měření také informace o kontextu, ve kterém byla tato data pořízena, protože to pak může ovlivnit následnou analýzu (viz. kapitola 2.3.2 a pojem *funkční definice* metriky - dohledatelnost).

Jako příklad tohoto tvrzení uvedeme jednoduchou situaci. Předpokládejme, že sledujeme časovou náročnost odstraňování chyb z kódu. Zaznamenáváme čas potřebný k nalezení a odstranění chyby z daného modulu. Pokud při analýze shledáme tento proces jako nestabilní a budeme se snažit najít příčiny takového chování, budeme jistě potřebovat ještě další informace vztahující se k měření. Zajímat nás bude především velikost modulu, která má přímý vliv na dobu potřebnou k odstranění nebo nalezení chyby. Další informace připojené k měření tak mohou velmi ulehčit hledání příčin nestability. Pokud tedy budeme chtít posuzovat proces jako celek, je vhodné mít také k dispozici údaje o jeho jednotlivých podprocesech, na základě kterých můžeme hlouběji a přesněji analyzovat celkové chování. Úsilí vynaložené na sběr těchto podpůrných dat by však v žádném případě nemělo

překročit únosné meze. S výhodou lze proto využít různé postupy, které generují potřebná data automaticky a ve velmi krátkém čase a nevyžadují tak lidskou práci navíc. V opačném případě bychom se totiž mohli setkat s odmítnutím metrik tak, jak to bylo popsáno v kapitole 2.4.1, kde byl zdůrazněn význam lidského faktoru.

Při kombinování různých výsledků měření však může vzniknout i opačná situace. Máme například k dispozici výsledky měření několika podprocesů, které však všechny vykazují známky nestability, a my budeme chtít analyzovat celkové chování procesu. Pokud takováto data nevhodně agregujeme, můžeme získat mylnou představu o tom, že celý proces je stabilní. Nevhodné sloučení několika měření totiž může zakrýt nežádoucí chování procesu. Při výpočtu kontrolních mezí dojde k jejich posunutí směrem dál od sebe a proces se tak bude mylně zdát celkově jako stabilní.

Dalším faktorem, který zásadně ovlivňuje sběr a analýzu dat, je frekvence s jakou jsou měření prováděna. Ta totiž přímo ovlivňuje množství nasbíraných dat a s tím související velikost skupin. Někdy povaha procesu dovoluje zvolit frekvenci vzorkování a velikost skupin libovolně, v jiných případech se musíme přizpůsobit procesu, který poskytuje data periodicky (reporty, měsíční vyúčtování, absence, denní výkazy hodin a další). Pro procesy, ve kterých lze volit frekvenci vzorkování dat a velikost skupin libovolně lze použít X-bar a R graf. Naopak u procesů, u kterých jsme omezeni, se budeme nejčastěji spoléhat na XmR graf.

Pokud tedy můžeme zvolit velikost skupiny libovolně, musíme brát ohled na první princip logického seskupování dat, který říká, že data ve skupině musejí být co nejvíce homogenní. Pokud zvolíme velikost skupiny nevhodně tak, že data v ní budou pokrývat příliš velké časové období, mohou se v nich vyskytnout různé skoky a posuny hodnot, což má za následek rozšíření kontrolních mezí grafů a ty se tak stávají méně citlivé a při odhalování nestability procesu méně nespolehlivé. Proces se totiž může zdánlivě zdát pod kontrolou, což ve skutečnosti nemusí být pravda. Pokud naopak zvolíme pro seskupování dat příliš malou skupinu, může nastat právě opačný případ, který je v rozporu s druhým principem logického seskupování. Průměrné hodnoty jednotlivých skupin mohou být jen velmi málo rozdílné, což má za následek velké zúžení kontrolních mezí a ty se tak pochopitelně stávají velmi citlivé.

Nehomogenní data ve skupinách tak negativně ovlivňují výpočet kontrolních mezí. Důležité je proto volit velikost skupin tak, aby následná analýza co nejlépe odpovídala na otázky, které chceme zodpovědět, nebo byla prováděna s cílem zvýšení kvality a vyspělosti procesu.

Hledání příčin nestability procesu je velmi náročná a zdoluhavá práce. Samotný proces se totiž skládá z mnoha částí (viz. kapitola 2 - definice procesu) a při identifikaci příčin nestability se proto musíme zaměřit na všechny tyto jeho části. Nejlepší je v tomto ohledu postupovat po malých krocích. Rozhodně není možné komplexně analyzovat chování celého procesu. Důležité je snažit se nalézt a izolovat co nejmenší části problémů, například pomocí metody analýzy nezvyklých vzorů chování (viz. následující kapitola). Musíme především sledovat vstupy a hledat zámky nestability právě v nich.

Dále se také musíme zabývat otázkou správného měření hodnot. Často totiž dochází v důsledku chybného odečítání nebo naopak zaokrouhlování k chybám, které se pak jen obtížně zpětně odhalují. Nevhodné zaokrouhlování může vést ke zkreslení kontrolních mezí grafů a analyzovaný proces se pak může jevit mylně jako stabilní nebo právě naopak. Pokud chceme měřit a zaokrouhlovat hodnoty spojitých veličin, je nutné zajistit, aby počet možných výsledků byl alespoň pět [FLO99]. Pouze v tomto případě budou všechny grafy poskytovat spolehlivé kontrolní meze.

Pokud chceme odstranit příčiny nestability procesu, musíme dbát na to, abychom proces nezměnili. Získali bychom tak vlastně nový proces, který by sice s tím původním měl spoustu společných vlastností, ale pro potřeby analýzy bychom museli získat nová data a na jejich základě posuzovat od začátku stabilitu procesu. Pokud se tomuto chceme vyhnout, musíme provádět takové kroky, které nebudou mít zásadní dopad na chování procesu a budeme tak moct využít již jednou

naměřená data. V podstatě se jedná o to, provádět pouze nápravná opatření a proces jako takový neměnit. Před i po provedení změn musí být proces v konzistentním stavu a metriky dosud používané musejí být i po provedení změn platné. Získáme tím také výhodu porovnání výkonnosti procesu před i po změně a díky tomu můžeme posoudit vhodnost a dopad provedených změn.

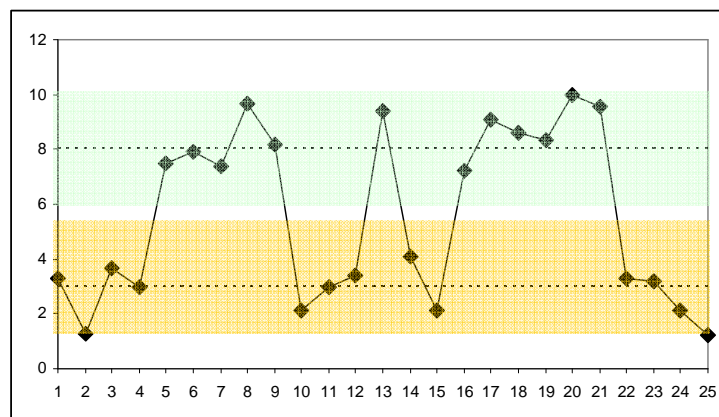
3.6.1 Sledování trendů

V následující podkapitole bude na základě několika ukázkových příkladů demonstrována další metoda analýzy grafů. Jedná se o identifikaci neobvyklých vzorů v datech. Nalezení takovýchto vzorů nám může pomoci odstranit příčiny nestability procesu a tato metoda se tak může stát dalším krokem při zvyšování kvality procesů.

V jedné z předchozích kapitol byla popsána metoda odhalování příčin nestability z grafů – čtyři testy nestability (kapitola 3.1.1). Toto však není jediný způsob, jak lze data analyzovat. Hledat pouze odlehle hodnoty není dostačující. Pro důkladnou analýzu je třeba se zaměřit také na rozložení dat v závislosti na čase jejich pořízení. Z tohoto důvodu hledáme v datech neobvyklé vzory chování, které mohou signalizovat nestabilitu procesu. Při sledování trendů grafů lze s výhodou použít XmR graf, ve kterém se trendy identifikují nejsnáze. Nyní si popíšeme nejčastější vzory.

3.6.1.1 Kompozice

V softwarových procesech se lze setkat s několika neobvyklými vzory. První z nich se nazývá kompozice a je charakterizován směsicí podobných hodnot, které mají tendenci se shlukovat okolo dvou a více navzájem různých středových linií. Na Obrázek 3-6 lze vidět příklad, kde data vykazují zvláštní charakter a mají tendenci se shlukovat okolo dvou středových linií (naznačeny tečkovaně).



Obrázek 3-6 Neobvyklé vzory chování: kompozice dvou procesů (zelená a žlutá)

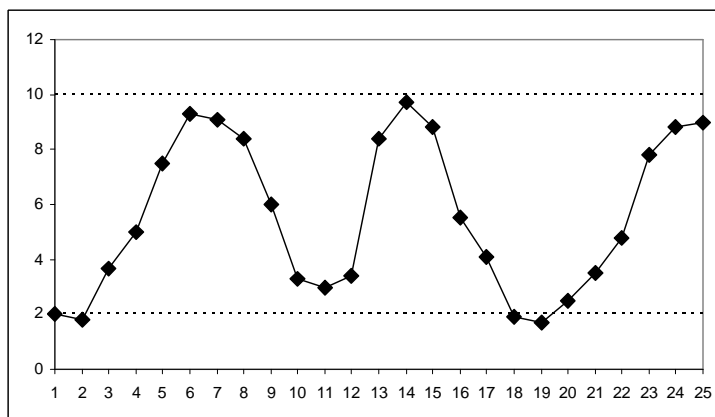
Nejčastější příčinou tohoto chování je situace, kdy se daný proces skládá z více podprocesů a každý z nich trvá různou dobu. Jako příklad uvedeme čas potřebný pro naprogramování jisté třídy, modulu nebo obecně nějakého chování. Je pochopitelné, že u rozsáhlých tříd, které realizují složité chování, bude tento čas značně vyšší. V grafu se pak tento vzor může projevit tak, že například u tříd o jednom tisících a více řádcích jsou příslušné hodnoty značně vyšší než u tříd menších.

Pokud se nám podaří vzor identifikovat, je vhodné data rozdělit do skupin podle nějakého kritéria (např. počet řádek kódu) nebo můžeme použít histogram hodnot, ze kterého lze jednoduše identifikovat počet a hranice shluků. Tento případ je vidět na Obrázek 3-6, kde jsou shluky vyznačeny zelenou a žlutou barvou. Identifikované shluky pak analyzujeme každý zvlášť. Tímto docílíme přesnější analýzy chování procesů.

3.6.1.2 Cyklus

Vzor cyklus je charakteristický periodickými změnami v datech. Ta pravidelně mění trend a toto chování se neustále opakuje v určitých cyklech. Takovéto chování je příčinou nestability procesu, kterou nemusíme být schopni odhalit pomocí výpočtu kontrolních mezí.

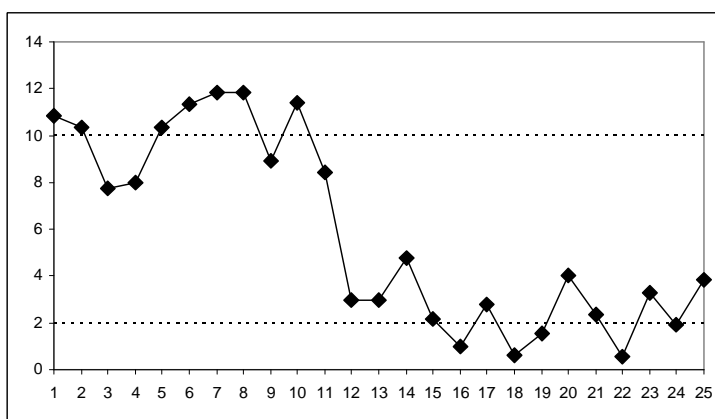
Toto chování se velmi často vyskytuje u metrik sledujících čas potřebný k vyřešení jistého úkolu a nejlépe jej odhalíme pomocí XmR, X-bar/R, C a U grafů [FLO99]. Délka periody může napovídat o příčinách nestability. Příklad vzory cyklus je uveden na následujícím obrázku.



Obrázek 3-7 Neobvyklé vzory chování: cyklus

3.6.1.3 Skok

Tento vzor chování procesu je typický náhlým odchýlením dat směrem nahoru nebo dolů od původních hodnot. Příčinou může být změna personálu. Například přijetí nového zaměstnance zefektivní práci celého týmu a čas potřebný na řešení jednotlivých úkolů značně klesne (skok směrem dolů). Naopak přijímání nové technologie vyžaduje určitý čas navíc, aby se s ní mohli seznámit všichni zaměstnanci. Tato situace pak může být v grafu charakterizována skokem nahoru.



Obrázek 3-8 Neobvyklé vzory chování: skok směrem dolů

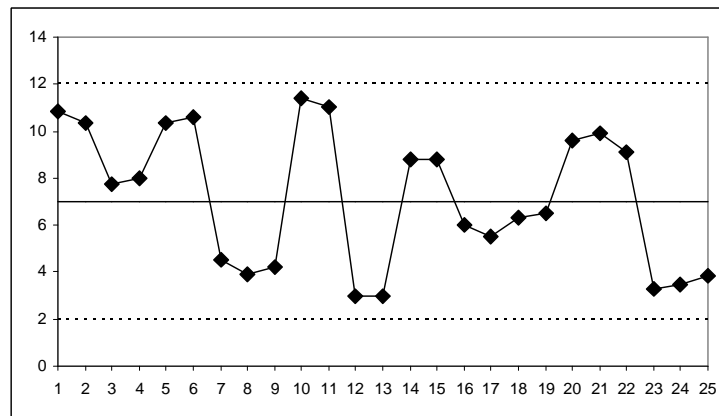
Výskyt vzoru skok u R grafů znamená změnu v kolísání procesu. Pokud se jedná o skok směrem nahoru, variabilita měřených dat uvnitř skupin se zvýšila a je nutné hledat příčiny takového chování. Naopak skok směrem dolů znamená jistou stabilizaci procesu a je tudíž žádoucí.

S takovými daty lze pracovat podobně jako tomu bylo u vzoru kompozice. Data lze rozdělit do dvou skupin. Jedna skupina bude obsahovat data, která byla naměřena před změnou, a druhá ta ostatní. Pro rozdělení lze opět použít histogram hodnot. Tyto dvě skupiny dat je pak možné podrobit zvláštní analýze.

3.6.1.4 Shlukování

Shlukování je dalším vzorem chování, který můžeme v datech pozorovat. Jeho příčinou jsou náhlé změny v procesu, které většinou trvají pouze 2 až 4 měření a v datech se projevují jako shluky velmi blízkých hodnot.

Jako příklad může sloužit vývoj softwarové aplikace, během kterého se úsilí vynaložené všemi členy týmu může měnit skokově po několika pracovních dnech. Důvody mohou být různé. Například nepřesná specifikace může na pár dnů zpomalit vývoj nebo naopak s blížícím se koncem etapy se může pracovní výkon dočasně zvýšit. Pokud takovýto vzor v datech identifikujeme a chceme odstranit jeho příčiny, vždy je nutné se zaměřit na jednotlivé skupiny dat. Musíme především analyzovat, co mají data ve skupinách společné a poté zkoumat rozdíly mezi jednotlivými skupinami.

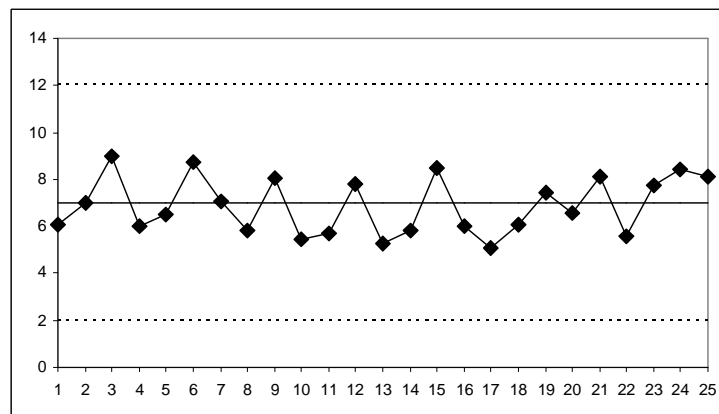


Obrázek 3-9 Neobvyklé vzory chování: shlukování

3.6.1.5 Stratifikace

Tento vzor je charakteristický tím, že všechna data jsou rozmístěna velmi blízko středové linie a jen velmi málo a nebo vůbec žádná z hodnot se blíží k horní nebo dolní mezi. Příklad takovýto dat je zachycen na Obrázek 3-10.

Příčinou tohoto chování není proces samotný, ale špatně nastavené měření a na jeho základě chybně vypočítané kontrolní meze. Například u X-bar grafů může k takovéto chybě dojít v okamžiku, kdy se do skupin jednotlivých měření dostanou silně nehomogenní data. Ty pak způsobí oddálení kontrolních mezí a proces se může mylně zdát jako stabilní. Naopak u R grafu může k tomuto chování dojít v důsledku jen velmi malých změn mezi jednotlivými skupinami. Stát se tak může opět při špatné volbě metrik a rozdělení dat do skupin.



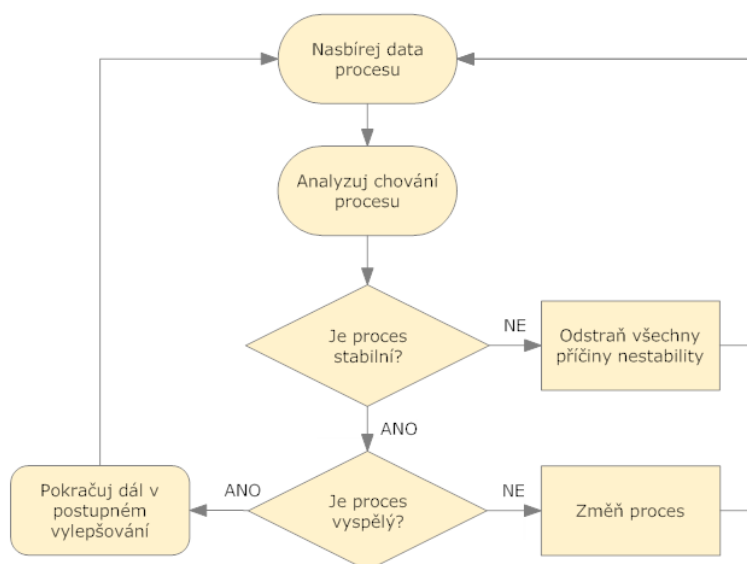
Obrázek 3-10 Neobvyklé vzory chování: stratifikace

Stratifikace tedy velmi úzce souvisí s definicí metriky a vzniká jako důsledek chybně zvolených měření, kvůli kterým jsou vypočítané kontrolní meze příliš vzdálené. Na tomto místě ještě jednou proto připomeňme dva základní principy logického seskupování dat. První z nich zdůrazňuje podobnost a homogenitu dat uvnitř všech skupin měření. Rozdíly by naopak měly být především mezi těmito skupinami.

3.6.2 Vyspělost procesu

Tato podkapitola přináší stručný přehled technik, které se uplatňují při zvyšování vyspělosti procesu (*process capability*). Navazuje volně na kapitolu 3.1.1 Vyhodnocování stability procesu a dále rozšiřuje poznatky v ní uvedené.

Vyspělost procesu lze definovat jako: „*Schopnost procesu plnit předem stanovené obchodní cíle*“ [FLO99]. Jinými slovy lze říct, že všechna měření procesu se musí pohybovat v předem stanovených a jasně specifikovaných hranicích, které nesmějí v žádném případě překročit. Znalost toho, jak moc je daný proces vyspělý, je důležitá především z hlediska predikce kvality jeho výstupů a možností jeho plánování. Zvyšování vyspělosti je tedy dalším krokem při zlepšování procesů. Před samotným posuzováním míry vyspělosti procesu musíme zajistit, aby byl proces pod statistickou kontrolou (viz. Obrázek 3-11 Schéma zlepšování procesů). Těto problematice byla věnována kapitola 3.1.1.



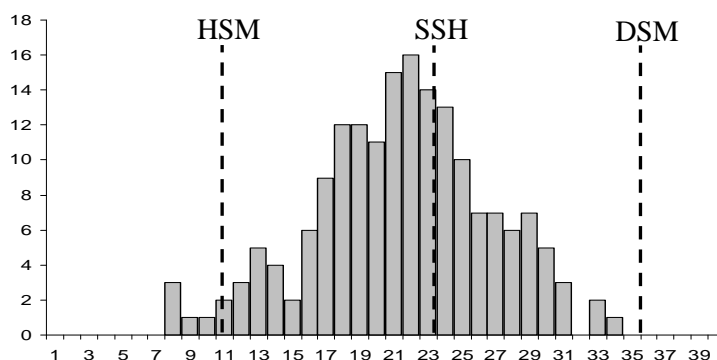
Obrázek 3-11 Schéma zlepšování procesů.

Při definování procesu je nutné do definice zahrnout i požadavky, které jsou na proces kladeny a které budou v budoucnu sloužit jako vodítko pro posuzování vyspělosti procesu s cílem jej zlepšit. Může se jednat například o požadavky na celkové náklady procesu, které lze vyjádřit určitou maximální hodnotou, kterou je nutné dodržet. Dále to mohou být různé limity a hranice, ve kterých se musí proces pohybovat, ať už se jedná například o časové hledisko nebo o požadavky na vstupy či výstupy procesu. V každém případě je nutné tyto hodnoty uvést do definice procesu přesně. Obvykle se definuje horní specifikační mez (HSM), která udává horní hranici chování procesu, dolní specifikační mez (DSM), která naopak udává spodní mez chování procesu, a pro úplnost lze specifikovat také střední hodnotu (SSH), která bude při posuzování vyspělosti procesu sloužit jako vodítko při určování centrální tendence chování procesu. Je pochopitelné, že v některých případech se neuvádějí všechny tři hodnoty, ale některé z nich se vypouštějí. Pokud například budeme zkoumat

počet chyb v kódu, nemusíme uvést dolní specifikační mez, která je pochopitelně rovna nule, protože se snažíme o naprosto bezchybný kód.

Musíme si tedy nejdříve při definici procesu ujasnit s jakým cílem chceme proces zlepšovat a definovat reálné a splnitelné limity. V opačném případě bychom totiž mohli proces přivést do nekonečné smyčky analyzování a zlepšování, které by však nemělo reálné výsledky.

Pokud výsledky provedených analýz na dostatečném množství dat (viz. kapitola 3.5.1 Hromadění dat) ukazují, že je daný proces pod statistickou kontrolou, lze přikročit k posuzování vyspělosti procesu. Tato fáze má dva kroky. V prvním kroku je nutné transformovat naměřená data nejlépe pomocí histogramu. Poté ve druhém kroku dojde k porovnání stanovených mezí s předem definovanými. Toto lze opět provést pomocí histogramu, do kterého zaneseme předem definované hodnoty. Situaci zachycuje Obrázek 3-12.



Obrázek 3-12 Histogram naměřených dat použitý pro posuzování míry vyspělosti procesu (HSM – horní specifikační mez, SSH – střední specifikační hodnota, DSM – dolní specifikační mez)

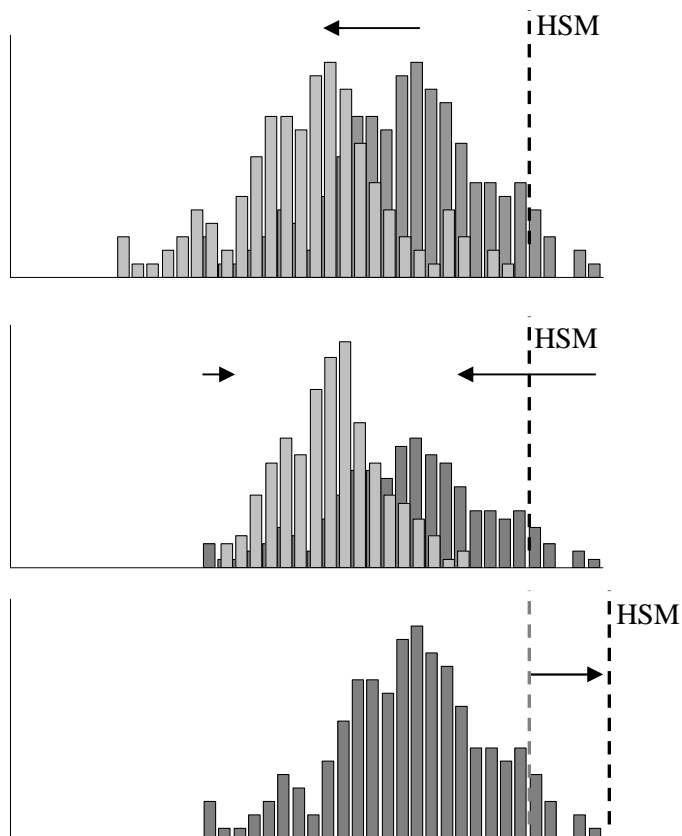
Pokud proces splňuje limity a předem stanovené požadavky a jeho chování je stabilní, jeho míra vyspělosti se v porovnání s předchozím stavem zvýšila. Pokud však proces vykazuje stabilní chování, avšak ne všechna jeho data spadají do stanovených mezí, je nutné provést nápravná opatření, která budou vést ke splnění všech požadavků. U všech procesů je tento krok tím nejobtížnějším při zlepšování procesů. Dvojnásob to pak platí u softwarových procesů. Jedná se o nalezení takových míst v procesu, díky nimž bude potenciálně možné zvýšit výkonnost procesu a toto nemusí být vždy snadné. Vždy závisí na konkrétní situaci a na konkrétním prostředí. Nelze proto podat jednoznačný návod, jak postupovat. Obecně je však nutné mít dobrou znalost daného procesu a při hledání možností jeho zlepšení lze také vyjít z definice procesu, která byla uvedena v kapitole 2 Management procesů. Je tedy nutné zkoumat především vstupy a výstupy procesu, lidské zdroje, nástroje, aktivity a potřebné materiály.

Existují tři základní postupu, které lze uplatnit při zvyšování výkonnosti procesu. Všechny graficky znázorňuje Obrázek 3-13, ve kterém je překročena horní specifikační mez.

Prvním z nich je posun střední hodnoty procesu (Obrázek 3-13 nahoře). Jedná se o situaci nejběžnější. Snahou je udělat v procesu takové úpravy, které budou vést ke změně jeho chování v jednom nebo ve druhém směru. Konkrétní kroky mohou znamenat například důkladné proškolení zaměstnanců, které bude mít za následek to, že svou práci budou vykonávat efektivněji, čímž se docílí například zkrácení doby potřebné pro vykonání procesu. Obdobnou situací je implementace nových a vyspělejších technologií do procesu. Při analýze procesu, u kterého chceme vychýlit jeho běžné chování, se musíme zaměřit především na data, která se pohybují okolo středních hodnot. V těchto datech musíme hledat společné atributy a hledat možnosti případných změn. Opět je vhodné se zaměřit na definici procesu uvedenou výše a z ní vycházet.

Na tomto místě je nutné poznamenat, že každá změna v procesu vlastně vytvoří úplně nový proces, s jehož měřením a analýzou musíme začít znovu od začátku. Všechna předchozí data nelze již dále používat a nelze z nich ani vycházet. Proto je fáze zvyšování vspělosti procesu jednou z časově nejnáročnějších. Vždy totiž musíme po provedení změn v procesu nasbírat dostatečné množství dat, ta analyzovat a určit, zda se proces chová stabilně. Až poté lze posuzovat vspělost procesu.

Druhou situací je snížení variability procesu (Obrázek 3-13 uprostřed). Při analýze se zaměřujeme především na odlehlá data a pokoušíme se je odstranit jednotlivými úpravami procesu. Typicky se může jednat o nastavení různých kontrol, které budou mít za cíl hlídat a případně korigovat příliš velkou variabilitu v jednom nebo ve druhém směru.



Obrázek 3-13 Postupy při zvyšování výkonnosti procesu. Horní graf ukazuje posun středové hodnoty, prostřední graf snížení variability a spodní graf posun specifikační meze.

Obdobně jako u předešlé situace lze také provést různá školení personálu, která budou mít za cíl zefektivnit práci. Konkrétní postup vždy záleží na dané situaci. Opět je však nutné poznamenat, že jakákoliv změna procesu vytváří proces úplně nový, u kterého je nutné nejdříve důkladně posoudit jeho stability a teprve poté lze přikročit ke zvyšování jeho vspělosti.

Třetí situace velice úzce souvisí s definicí procesu, ve které je mimo jiné nutné stanovit specifikační meze a pokud jsou nastaveny nevhodně, může to vést k mnoha problémům. Tato situace nastává velmi často u nových procesů, které definujeme bez předchozí zkušenosti a specifikační meze pouze odhadujeme. Pokud se v praxi ukáže, že proces nemůže specifikacím vyhovět, je nutné zvážit jejich uvolnění. Tento krok se může zdát sporný. Posunutím mezí totiž proces v žádném ohledu nezlepšíme. Vytvoříme však nové specifikační meze, o kterých budeme mít jistotu, že jsou dosažitelné a položíme tak pevný základ pro další zlepšování procesu.

Celá problematika zvyšování míry vspělosti procesů je složitější. Tato podkapitola přinesla stručný přehled toho, jak se posuzuje míra vspělosti procesu jakožto další krok jeho zlepšování a doplnila tak již dříve zmíněné údaje.

4 Návrh řešení

Tato kapitola podrobně popisuje návrh aplikace, která je součástí diplomové práce. Popisuje základní teoretická východiska dané problematiky, rozebírá nejdůležitější myšlenky aplikace a zároveň shrnuje její možnosti s důrazem na vlastní přínos.

Aplikace se svou funkcionalitou zaměřuje na poslední fázi měření a vylepšování procesů, tedy na analýzu naměřených dat. Koncept aplikace je postaven na výše popsané metodice statistického přístupu k analýze chování procesu. Zahrnuje všechny doporučené postupy a přináší řadu vylepšení, která pomáhají proces analýzy dat zjednodušit a co nejvíce kroků automatizovat.

4.1 Koncepce aplikace

Řešení je navrženo především s důrazem na generičnost a obecnou použitelnost. Pomocí aplikace lze tedy analyzovat libovolný proces, nemusí se jednat pouze o softwarové procesy. Tato univerzálnost je dále podpořena velmi jednoduchým databázovým rozhraním, na které nejsou kladena prakticky žádná omezení nebo nároky. Podrobněji o tomto tématu bude řeč v kapitole 5.

Během fáze návrhu aplikace byly stanoveny následující hlavní cíle:

- zachovat koncept aplikace co nejvíce generický
- práci uživateli co nejvíce usnadnit a automatizovat
- vytvořit pokročilé uživatelské rozhraní

Nyní k jednotlivým bodům podrobněji. Jako první cíl bylo stanoveno navrhnout a implementovat aplikaci s důrazem na obecnou použitelnost. Výhodou tohoto je, že lze pomocí aplikace analyzovat libovolný proces. Snahou bylo především zjednodušit práci s daty do té míry, že budou naprosto nezávislá na použité platformě a jejich definice nebude ničím vázána. Z toho důvodu byla jako zdroj dat zvolena databáze. Definice metriky pak představuje jednu tabulku této databáze a atributy metriky jsou vlastně sloupce dané tabulky. Výhodou tohoto přístupu je také možnost zadávat data do databáze prakticky z libovolného místa a z libovolného prostředí, například pomocí webového rozhraní, speciálního klienta nebo pomocí libovolné jiné aplikace. Aplikaci je tak zároveň možné připojit na libovolnou databázi a analyzovat libovolná data. Generičnost řešení spočívá také v tom, že lze definovat libovolný počet analýz a ty provádět paralelně. Díky tomu je možné například porovnávat několik paralelních procesů zároveň v jednom okně nebo analyzovat chování procesu z různých úhlů pohledu. Tohoto využijeme například v situaci, kdy budeme chtít porovnat výkonnost jednotlivých členů týmu nebo porovnat výsledky několika analýz. Každé analýze lze navíc přiřadit libovolný typ grafu a nastavit všechny parametry nezávisle na ostatních definovaných analýzách. Poznamenejme, že v kontextu aplikace hovoříme o analýze jako o skupině grafů, které poskytují různé pohledy na stejná data. Tímto se rozumí například dvojice X-Bar/R graf, které zobrazují dva pohledy na ta samá data avšak hodnoty obou grafů jsou vypočítávány odlišným způsobem (viz. kapitola 3.2 Grafy chování procesů). Proto lze v aplikaci definovat libovolné množství analýz, jelikož je tento požadavek chápán jako zobrazení více grafů. Více o možnostech definic analýz v následující kapitole nebo v ukázkových videích, které jsou na příloženém DVD nosiči (viz. kapitola 6.1.3).

Na druhou stranu je však nutné přiznat, že generičnost celého řešení klade nemalé nároky na uživatele. Ten musí být obeznámen s danou problematikou a musí přesně vědět, čeho chce analýzou dosáhnout. Je také nutné znát definice metrik a rozumět analyzovaným procesům, což je ovšem obecný předpoklad každé takové analýzy.

K tomu, aby uživatel mohl s aplikací pracovat pohodlně a nemusel se zabývat příliš velkým množstvím detailů, bylo do návrhu aplikace zahrnuto také několik vylepšení, která usnadňují uživateli práci a automatizují některé kroky. Jedná se o automatické nastavení parametrů analýzy pomocí předem definovaných modelových situací. Pokud si uživatel není jistý, jaký typ grafu pro danou analýzu zvolit nebo neví, jak nastavit ostatní atributy, může toto za něj provést aplikace automaticky. Je nutné pouze zvolit analýzu, kterou chceme nastavit, a přiřadit jí modelovou situaci, která nejvíce odpovídá požadavkům. Aplikace automaticky doplní všechny potřebné údaje. Více o této možnosti v demonstračních videích, které jsou na přiloženém DVD nosiči.

V aplikaci je dále zabudován systém automatické kontroly definic metrik, tzv. *checker*. Ten při každé změně definice metriky ověří její platnost a v případě jakékoliv chyby vypíše příslušné chybové hlášky. Uživatel má tak možnost definici podle pokynů opravit. Zároveň nemusí hlídat všechny závislosti nastavovaných parametrů, aplikace toto dělá za něj. Zabrání se tím také vzniku chyb při analýze. O automatické kontrole definic analýz je také možné se dozvědět více v demonstračních videích na přiloženém DVD.

Aplikace bude rozdělena na dvě hlavní části. V té první bude uživatel specifikovat analýzu, kterou si přeje provést a druhá část bude sloužit pro samotné provádění analýzy. První část bude zahrnovat výběr metriky, což ve své podstatě bude odpovídat výběru tabulky z databáze. Dále zde bude možné na tato zvolená data nahlížet. Hlavním důvodem zobrazení dat uživateli je snadnější orientace při definici analýz, což bude hlavní krok v první části aplikace. Uživatel zde na základě vlastních požadavků bude specifikovat typ a počet analýz, které si přeje v souvislosti se zvolenou metrikou provést. Jeho úkolem zde bude vytvořit libovolný počet analýz a nastavit jim všechny potřebné parametry. Jak již bylo jednou zmíněno, jedná se o krok, k jehož provedení musí mít uživatel potřebné znalosti a případně i zkušenosti. Z tohoto důvodu bude aplikace obsahovat modul pro automatické nastavování parametrů analýz a zároveň také komponentu (*checker*), která bude automaticky a zcela samostatně ověřovat správnost definovaných analýz a v případě výskytu chyb o tomto bude informovat uživatele. Ten tak bude mít možnost všechny případné chyby odstranit a zaměřit se především na samotný cíl analýz. V okamžiku, kdy budou všechny analýzy specifikované správně, uživateli bude umožněno přejít do další části aplikace.

Druhá část aplikace tak bude sloužit pro práci s grafy a vykonávání samotné analýzy a jejich jednotlivých kroků. Jednou z nejdůležitějších předností aplikace by mělo být přehledné a pohodlné uživatelské rozhraní, které bude zohledněno především v této fázi práce s aplikací. Uživatel tak bude mít možnost provádět všechny kroky intuitivně. Do návrhu proto byly začleněny pokročilé techniky práce s grafy, kterým se mimo jiné podrobněji věnuje následující kapitola. Uživatel bude mít možnost v této fázi specifikovat jednotlivé skupiny dat pomocí výběr kurzorem myši, dále pomocí histogramu a nebo v kterémkoliv jiném grafu příslušné analýzy. Aplikace také bude obsahovat možnost automatického výběru skupin podle uživatelem zadaných parametrů, což opět pomůže analýzu dat usnadnit a některé opakující se kroky automatizovat. Hlavním účelem druhé části aplikace tedy bude vytvářet uživatelem definované skupiny dat v jednotlivých grafech analýz a na základě těchto definic pak nechat aplikaci automatiky vygenerovat nové iterace (viz. kapitola 3.5.2 Metoda eliminace odlehlých dat). Více o možnostech vytváření nových iterací (kroků analýz) se dozvíte v následující kapitole. V souvislosti s vytvářením skupin dat a prováděním analýzy je nutné zajistit uživateli možnost nahlížet na původní data, ze kterých jsou vypočítávány jednotlivé hodnoty v grafech. Toto bude zahrnovat možnost nechat si zobrazit všechna data a vhodné by také bylo ponechat uživateli možnost nechat si zobrazit pouze jím vybrané konkrétní hodnoty. Toho lze pak využít například v situaci, kdy bude zvolená analýza obsahovat velké množství dat nebo v případě, kdy bude orientace v datech komplikovanější. Uživatel bude mít tedy možnost zobrazit si data příslušné analýzy a navíc také z těchto dat bude mít možnost volit a nastavovat skupiny. Tímto se úroveň práce s aplikací opět posune výše.

Pokud tedy shrneme výše uvedené. Návrh aplikace respektuje metodiku statistického přístupu k analýze chování procesů a rozšiřuje ji ve třech hlavních směrech – generičnost řešení, maximální míra automatizace, propracované uživatelské rozhraní. Aplikace bude rozdělena na dvě hlavní části, první sloužící pro definici analýzy a druhá pro práci s grafy.

4.2 Diagramy případů použití

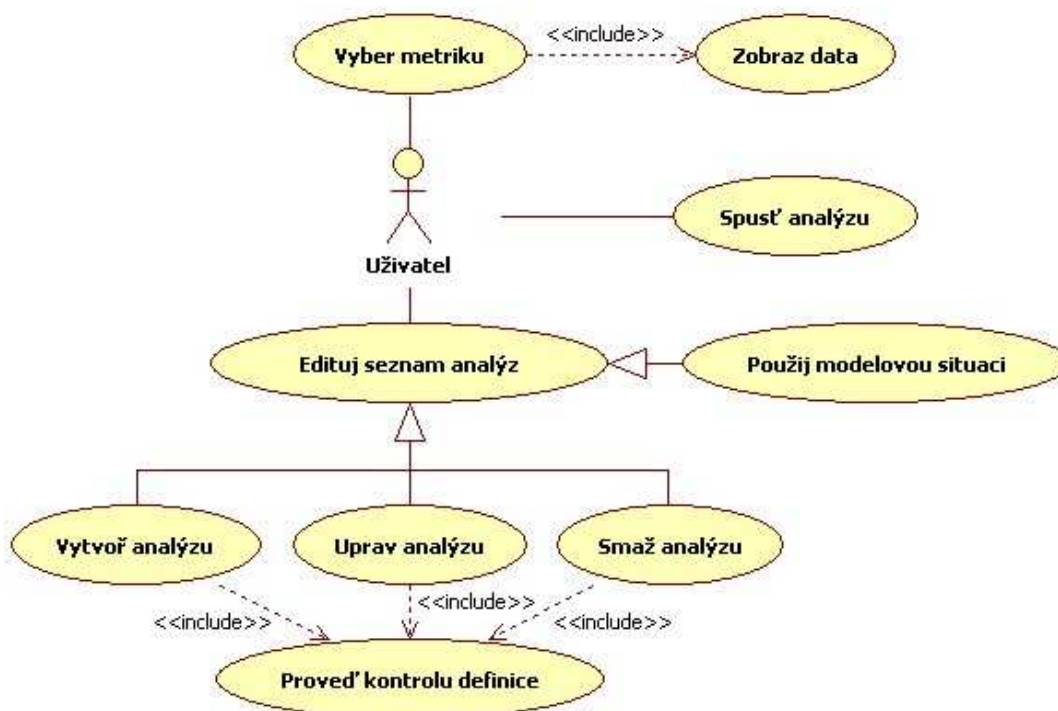
Nyní podrobněji rozebereme některé hlavní případy použití a detailně popíšeme hlavní myšlenky interakce uživatele se systémem. K tomuto účelu budou sloužit diagramy případů použití.

Diagram případů použití na Obrázek 4-1 zachycuje hlavního a zároveň jediného aktéra, kterým je *Uživatel*, a hlavní případy použití systému, které spouští. Tento diagram zachycuje aktéra v první části aplikace, která, jak již bylo zmíněno v předchozí podkapitole, bude sloužit pro výběr metrik a definici analýz.

Prvním případem, který uživatel bude spouštět, je „*Vyber metriku*“. V souladu s výše uvedeným návrhem bude tento krok znamenat požadavek aktéra na volbu tabulky, která obsahuje všechna data příslušné metriky. Aplikace načte a zobrazí tato data do vyhrazené části okna. Z tohoto důvodu také případ zahrnuje automatické vykonání případu *Zobraz data*.

Dalším krokem aktéra bude vytvořit seznam analýz, které chce na základě zvolené metriky provést (*Edituj seznam analýz*). Připomeňme, že v terminologii, kterou používá aplikace, se analýzou myslí skupina grafů, které poskytují různé pohledy na stejná data (např. dvojice X-Bar/R graf). Tento krok zahrnuje vytváření analýz (*Vytvoř analýzu*), mazání analýz (*Smaž analýzu*), které uživatel zadá chybně nebo se rozhodne je neprovádět, a v neposlední řadě také editace a úpravy jež definovaných analýz (*Uprav analýzu*). Při každé úpravě v seznamu analýz dojde k automatickému spuštění komponenty, která ověřuje správnost definice analýz (*Proved' kontrolu analýz*). Poslední možností bude použití modelové situace pro nastavení parametrů analýzy (*Použij modelovou situaci*).

Ve chvíli, kdy *Uživatel* ukončí zadávání seznamu analýz, přejde do druhé části aplikace (*Spust' analýzu*).

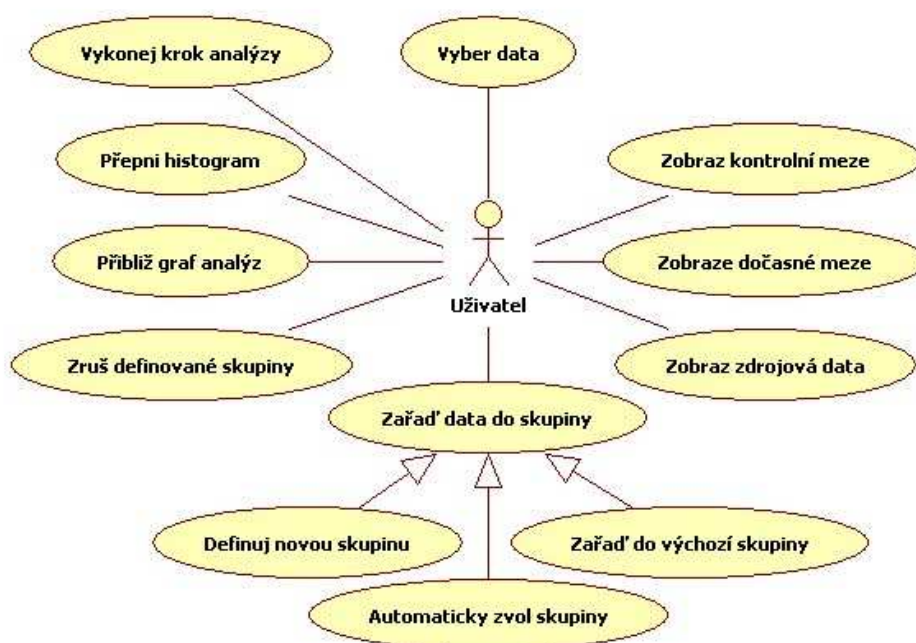


Obrázek 4-1 Diagram případů použití – fáze definice analýz

Ve druhé části aplikace, ve které se provádí samotná analýza, uživatel pracuje výhradně s grafy. Obrázek 4-2 zachycuje jednotlivé případy použití, které bude mít možnost *Uživatel* spouštět.

Jedním z hlavních případů použití je *Vyber data*. Uživatel bude vybírat pomocí několika výše popsaných způsobů data, se kterými bude následně pracovat. Například si pro ně může nechat spočítat dočasné kontrolní meze (*Zobraz dočasné meze*), zobrazit příslušná data (*Zobraz zdrojová data*) a také vybrané body bude moci zařadit do skupiny (*Zařad' data do skupiny*). Tento případ je zobecněním tří případů použití, které představují zařazení dat do nově definované skupiny (*Definuj novou skupinu*) nebo do skupiny výchozí (*Zařad' do výchozí skupiny*), která bude sdružovat všechny nepřirazené body grafu. V poslední variantě bude možné data zařazovat do skupin automaticky (*Automaticky zvol skupiny*) a to na základě zvolených prahových hodnot.

Možnost přiřadit data do výchozí skupiny (*Zařad' do výchozí skupiny*) byla dodatečně implementována během vývoje prvních verzí aplikace, jelikož předběžné výsledky práce s aplikací ukázaly, že je běžné během analýzy provádět změny v definovaných skupinách a jednotlivé body přidávat a také odebírat. Rozšířením tohoto případu je pak možnost zcela vymazat všechny definované skupiny dat (*Zruš definované skupiny*).



Obrázek 4-2 Diagram případů použití - fáze provádění analýzy

Základním přístupem při analýze dat pomocí statistických metod je postupné rozdělování dat do skupin podle daných kritérií a následné vytvoření nových grafů. Kritéria lze určit například pomocí kontrolních mezí. Po jejich zobrazení (*Zobraz kontrolní meze*) bude moci uživatel vizuálně posoudit pozici jednotlivých bodů v grafu vzhledem k těmto liniím a na základě toho zvolit skupiny. Pro každou definovanou skupinu dat se tak vytváří nový graf, který obsahuje právě data z dané skupiny. Ke všem nově vzniklým grafům se také vypočítají nové kontrolní meze a analýza se opakuje. Toto vytváření nových grafů je zahrnuto pod případem *Vykonej krok analýzy*.

Pro pohodlnou práci s grafy bude možné také možné zobrazit nebo skrýt histogram hodnot (*Přepni histogram*) a případně také přiblížit nebo zvětšit graf analýzy (*Přiblíž graf analýz*).

Výčet možností aplikace není jistě kompletní. Analýza a návrh řešení však odhalily základní požadavky na systém pro analýzu chování procesů, který bude dále rozšiřován podle potřeb během fáze implementace, které se věnuje následující kapitola.

5 Popis implementace

Tato kapitola podrobněji popisuje implementační stránku aplikace. Slouží také jako stručný uživatelský manuál, jelikož přináší přehled všech základních funkcí. Aplikace je rozdělena do dvou hlavních částí, kterým bude věnována pozornost. Cílem implementace bylo splnit třetí bod zadání této diplomové práce - implementovat prototyp aplikace. V závěru je jedna podkapitola věnována podrobnějšímu popisu konkrétní části aplikace.

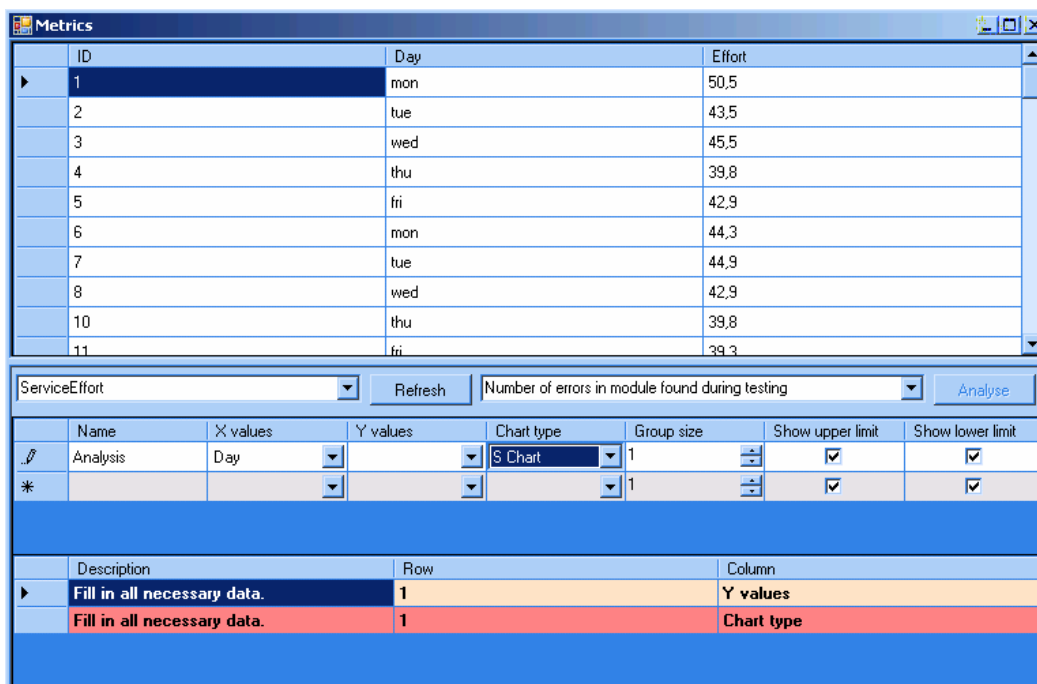
5.1 Základní rozdělení aplikace

Aplikace je vyvíjena na platformě Microsoft .NET 3.5 v jazyce C# v prostředí Microsoft Visual Studio 2008. Důvodů této volby je několik. Především jsou to mé bohaté zkušenosti s vývojem aplikací v jazyce C#, dále široké možnosti, které platforma nabízí, a v neposlední řadě volbu ovlivnily použité knihovny, které jsou vytvořené právě pro .NET. Jako zdroj dat slouží databáze Microsoft SQL, která obsahuje několik předem definovaných metrik.

Pro spuštění aplikace jsou požadovány následující komponenty:

- Microsoft .NET 3.5 SP1 Framework (www.microsoft.com/.NET/)
- Microsoft Chart Controls (code.msdn.microsoft.com/mschart)
- Microsoft Crystal Reports (www.sap.com/solutions/sapbusinessobjects/sme/reporting/crystalreports)
- Microsoft Data Access Application Block (<http://msdn.microsoft.com/en-us/library/cc309504.aspx>)

Aplikace je rozdělena na dvě hlavní části. V té první uživatel volí data, která chce analyzovat a nastavuje parametry analýzy. Druhá část pak zahrnuje samotnou analýzu a především práci s grafy. Nyní se budeme věnovat jednotlivým částem podrobněji.



Obrázek 5-1 První okno aplikace zahrnuje výběr dat a definici metrik.

První část aplikace slouží uživateli pro nastavení analýzy (viz. Obrázek 5-1). Po stisku tlačítka „Refresh“ aplikace načte všechny definované metriky (tabulky) z databáze a nabídne jejich volbu v příslušné rozvírací roletce. Pro snadnou analýzu by databáze měla obsahovat mimo jiné i sloupce, které budou každé měření blíže specifikovat (viz. kapitola 2.3.2 – dohledatelnost). Jedná se například o datum a čas pořízení záznamu, osobu zodpovědnou za pořízený záznam, popis a případně další užitečné informace vztahující se ke každému konkrétnímu záznamu. Tyto sloupce pak budou sloužit při analýze dat, která bude popsána dále.

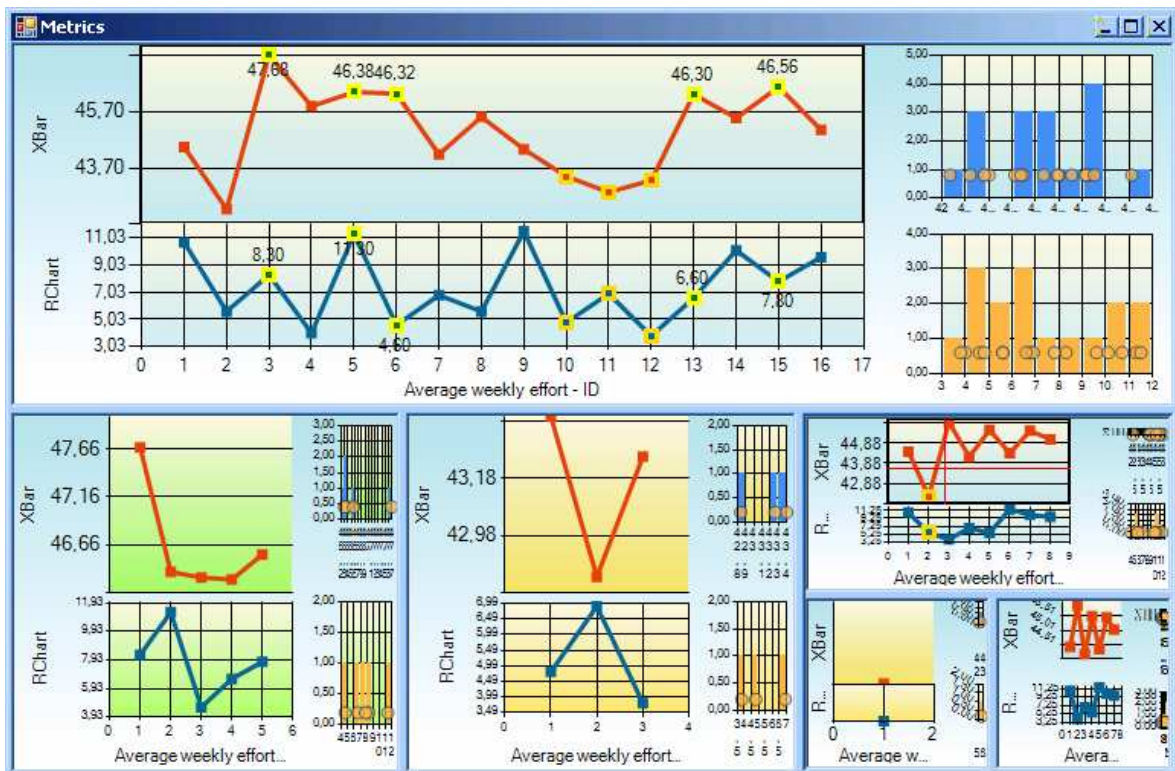
Uživatel následně vybere metriku (tabulku s daty), kterou chce analyzovat. V horní části okna se data příslušející zvolené metrice zobrazí a uživatel má možnost si je prohlížet. Poté je nutné definovat analýzu. K tomuto účelu slouží prostřední část okna, která obsahuje tabulku, jejíž jednotlivé řádky odpovídají jednotlivým analýzám. Ty lze libovolně přidávat, odebírat a měnit. Je nutné nastavit jméno analýzy, které slouží pro snadnější orientaci v grafech, zvolit sloupec tabulky (atribut metriky), který se stane zdrojem dat pro osu X a sloupec, ze kterého aplikace bude vypočítávat hodnoty na ose Y. Dále uživatel zvolí typ grafu, který chce použít pro analýzu, a specifikuje velikost skupiny v souladu s výše uvedenou teorií. Pro úplnost je možné pro každou definovanou metriku vypnout zobrazení horní nebo i dolní kontrolní meze, pokud si to uživatel přeje. Aplikace pak v náhledu grafu příslušné meze nezobrazuje. O možnostech definic analýz se lze dozvědět více v ukázkách na DVD.

Pro ulehčení práce a zjednodušení definic metrik byla do aplikace implementována možnost automatického nastavení metriky podle předem definovaných modelových situací. Tuto volbu je možné použít například v situaci, kdy si uživatel není jistý přesným nastavením parametrů analýzy. Jednoduše zvolí libovolnou definici analýzy tak, že označí příslušný řádek v tabulce definic kliknutím na jeho záhlaví a poté z roletky, která se nachází mezi tlačítky Refresh a Analyse, vybere typ analýzy, která nejvíce odpovídá jeho požadavkům. Aplikace pak automaticky doplní příslušné parametry. Přednastavené situace zahrnují výhradně analýzu softwarových procesů. Podrobný popis této možnosti opět zahrnují ukázková videa na DVD.

Implementace aplikace zahrnuje také tzv. *checker*. Jedná se o komponentu, která automaticky ověřuje platnost definic analýz a o všech případných chybách uživatele informuje v příslušném okně (viz. Obrázek 5-1 dole – dva červené řádky) a nabízí mu také informace o tom, jak tyto chyby odstranit. V okamžiku, kdy jsou definice všech analýz správné, je aktivováno tlačítko Analyse, kterým se přechází do druhé části aplikace.

Ještě před popisem druhé části aplikace, kterou je analýza pomocí grafů, je vhodné vysvětlit několik základních pojmů, které s ní souvisejí. K tomuto účelu bude sloužit vzorový příklad jednoduché analýzy z Obrázek 5-2 a také z Obrázek 5-5. Na něm je zachycen případ jedné studie dat, která se nazývá *Šetření (Investigation)*. Tento pojem zahrnuje všechny grafy v okně. Každé *šetření* má několik *kroků (Investigation step)*. Ty představují horizontálně oddělené skupiny grafů (např. horní nebo dolní polovina okna). Každé šetření obsahuje několik *analýz (Analysis)*, které odpovídají definicím z první části aplikace. Každá *analýza* může mít několik typů grafů a tyto se navzájem synchronizují při výběru hodnot. Ke každému grafu *analýzy* je možné nechat si zobrazit histogram příslušných hodnot, z něhož je možné opět data vybírat. Více o možnostech pokročilé synchronizace v následující kapitole a také v ukázkových videích na DVD, které je součástí této práce. Jednotlivým názvům odpovídá vždy jedna konkrétní třída v programu. Popis těchto tříd lze najít v následující podkapitole.

Při spuštění analýzy se vytvoří nové *šetření* a zároveň se vytvoří první *krok šetření (Investigation step)*, který obsahuje všechny definované *analýzy*. Základním principem analýzy je postupné rozdělování dat grafů do *skupin (Groups)* podle uživatelem definovaných kritérií (viz. kapitola 3.5.2) a následné automatické vytvoření nového *kroku šetření*, který obsahuje jednotlivé *kroky analýz* podle rozdělení do skupin.



Obrázek 5-2 Druhá část aplikace - analýza dat.

Postup analýzy dat je tedy následující. Uživatel nejdříve rozdělí data v prvním kroku *šetření* do skupin a to tak, že je vybere pomocí cursoru myši v grafu nebo pomocí histogramu. Označené body se v grafu zvýrazní a stejně tak se zvýrazní odpovídající si body ve všech grafech daného kroku analýzy. Pak pravým tlačítkem vyvolá kontextové menu a z něj zvolí položku *Add to group* a vybere, případně definuje, skupinu, do které chce data zařadit. Jednotlivé body grafu se barevně odlišují podle přiřazení do skupiny. Takto uživatel může definovat libovolný počet skupin a v okamžiku, kdy jsou data rozdělena, může dojít k automatickému vytvoření nového kroku *šetření*. Jen pro úplnost je dobré zmínit, že ne všechny body grafu musejí být přiřazeny do skupiny. Aplikace automaticky přiřadí všechny nedefinované body do výchozí skupiny. Nový krok *šetření* se vytvoří při volbě položky *Exekute step* z kontextového menu aplikace. Aplikace automaticky vytvoří sérii nových grafů, z nichž každý obsahuje data z jedné definované skupiny. Takto vytvořené kroky analýz lze iterativně dále analyzovat a znovu na základě definic skupin vytvářet kroky *šetření*.

Uživatel si dále může nechat zobrazit ke každému grafu kontrolní meze, které aplikace na základě definice analýz vypočítá. Zobrazování se provádí pomocí volby *Toggle limits* z kontextového menu. Přepínat lze mezi pěti režimy zobrazení kontrolních mezí. Více v ukázkových videích na DVD. Aplikace také umožňuje uživateli zobrazení tzv. dočasných kontrolních mezí, které jsou vypočítány z předem označených bodů. Takto si uživatel může nechat zobrazit meze jen pro určitou skupinu dat, aniž by musel vytvářet nový krok analýzy. Jedná se o další možnost, která usnadňuje práci s aplikací a posouvá možnosti analýzy.

Tímto výčet možností uživatelského rozhraní nekončí. Grafy lze také libovolně přibližovat případně dočasně vyjmout z analýzy a pracovat s nimi ve zvláštním okně, lze také měnit pořadí analýz, automaticky provádět kroky analýz a další. O všech možnostech, které aplikace nabízí se lze dozvědět z demonstračních videí, které jsou součástí příloženého DVD. Jejich podrobný výčet je popsán v následující kapitole.

5.2 Logická vrstva aplikace

Tato podkapitola se věnuje popisu části logické vrstvy aplikace, která je základním prvkem celého konceptu, čímž zároveň doplňuje informace uvedené v předchozí kapitole.

Logická vrstva aplikace je tvořena několika třídami, jejichž úkolem je zajistit načítání dat z databáze, jejich transformaci na informace, které jsou zobrazovány uživateli. Vrstva také zpracovává požadavky, které přicházejí směrem z presentační vrstvy od uživatele. Obrázek 5-3 zachycuje nejdůležitější třídy logické vrstvy a zobrazuje jejich vzájemné vztahy.

Nyní se zaměříme na popis jednotlivých tříd. Hlavní z nich je zachycena v diagramu nahoře vlevo. Jedná se o třídu **Investigation**, která je společná pro všechny grafy dané analýzy. Jejím hlavním účelem je poskytnout data potřebná pro výpočet hodnot jednotlivých grafů, k čemuž je určená proměnná *dataTable*. S každým voláním metody *AddAnalysis()* se ve třídě vytvoří nová logika analýzy, která je uložena do pole *logicMap*.

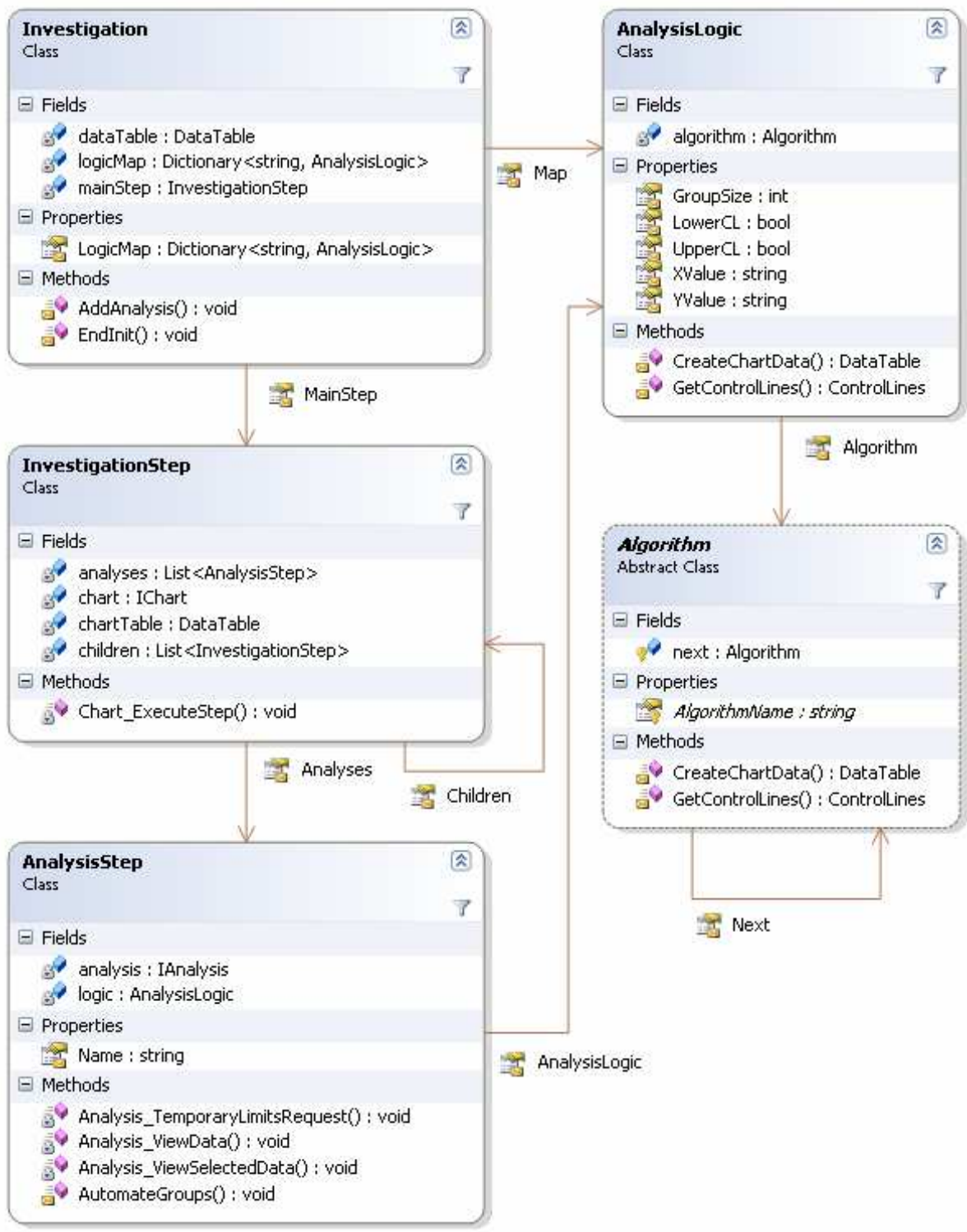
Logiku analýzy představuje třída **AnalysisLogic**. Ta v podstatě koresponduje s definicí analýzy z první části aplikace, ve které se zadávají jednotlivé její parametry. Mezi ně patří definice názvu dat pro osu X (*XValue*), definice názvu dat pro osu Y (*YValue*), zobrazení horních (*UpperCL*) případně spodních (*LowerCL*) kontrolních mezí a velikost skupiny (*GroupSize*). Třída pak na základě této definice poskytuje rozhraní pro generování dat k jednotlivým grafům pomocí volání metody *CreateChartData()* a pro zpracování požadavků na výpočet kontrolních mezí prostřednictvím metody *GetControlLines()*.

O tyto výpočty se stará třída **Algorithm**. Jedná se o abstraktní třídu, jejíž jednotlivé implementace představují konkrétní algoritmy výpočtu, například hodnot X-Bar grafu, R grafu a všech ostatních. Podstatnou vlastností každého algoritmu je proměnná *next*, která odkazuje na další algoritmus a tak vytváří jednosměrně vázaný seznam všech algoritmů, které jsou pro výpočet dat dané analýzy potřeba. Například kombinaci dvou grafů X-Bar/R odpovídá dvojice navzájem propojených algoritmů podle Obrázek 5-4. Toto řešení umožňuje vytvářet libovolné kombinace grafů dané analýzy bez nutnosti zásadně měnit koncept celé aplikace. Další předností je také snadná rozšiřovatelnost aplikace v případě, pokud zjistíme, že chceme k analýze použít vlastní graf a nebo změnit způsob výpočtu.

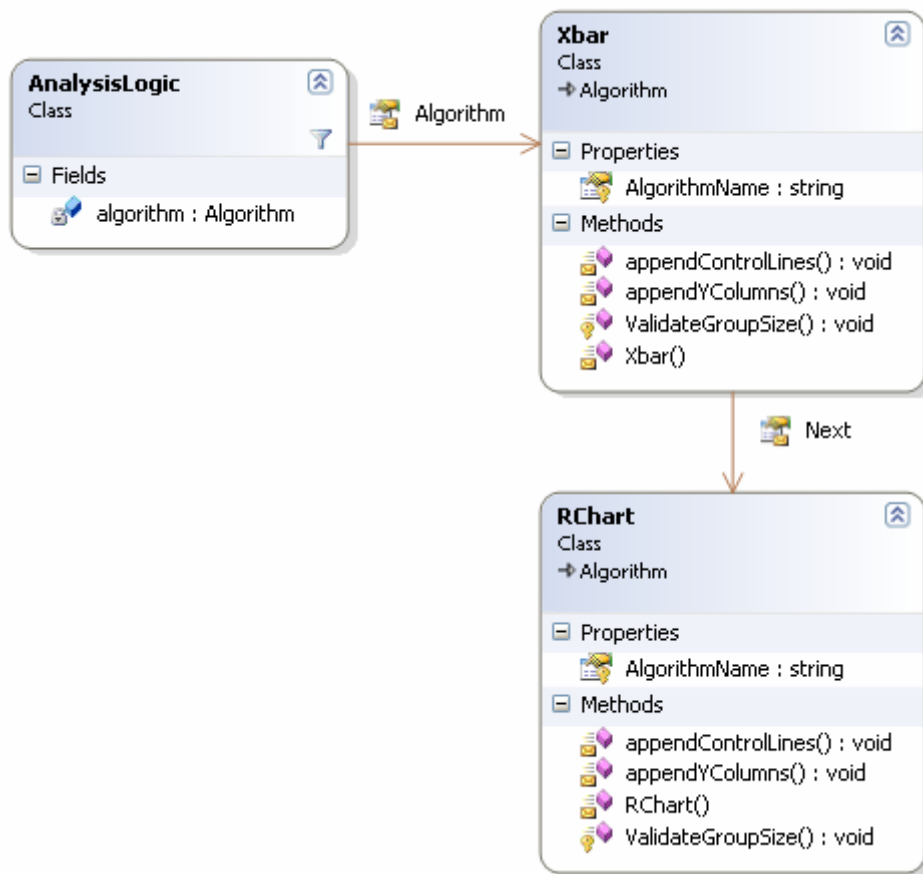
Nyní se vrátíme k popisu třídy **Investigation**, od kterého jsme odbočili. Série volání metod *AddAnalysis()* je zakončena voláním metody *EndInit()*, která automaticky vytvoří první krok analýzy v podobě třídy **InvestigationStep** a uloží jej do proměnné *mainStep*.

Hlavním úkolem třídy **InvestigationStep** je udržovat seznam všech analýz v proměnné *analyse* a odkaz na graf z presentační vrstvy v proměnné *chart*. V proměnné *chartTable* jsou uložena data vypočítaná pomocí logik jednotlivých analýz (**AnalysisLogic**). Tato data se pak zasílají grafům, od kterých třída získává a zpracovává požadavek na vytvoření nového kroku analýzy (**InvestigationStep**). Zpracování obstarává metoda *Chart_ExecuteStep()*, která přidá všechny nově vytvořené kroky šetření do seznamu *children*.

Poslední třídou je **AnalysisStep**. Ta koresponduje s jedním krokem analýzy, což je vlastně jeden nebo více vzájemně synchronizovaných grafů. Krok analýzy, který má svůj název uložen v proměnné *Name*, obsahuje odkaz *logic* na logiku (**AnalysisLogic**), pomocí které analýza transformuje zdrojová data na data jednotlivých grafů. Třída také uchovává odkaz *analysis* na odpovídající rozhraní dané presentační vrstvy. Prostřednictvím tohoto rozhraní třída dostává požadavky na výpočet dočasných kontrolních mezí (*Analysis_TemporaryLimitsRequest()*), požadavky na zobrazení dat (*Analysis_ViewData()* a *Analysis_ViewSelectedData()*) a na závěr také umožňuje automatické přiřazení bodů do skupin (*AutomateGroups()*).

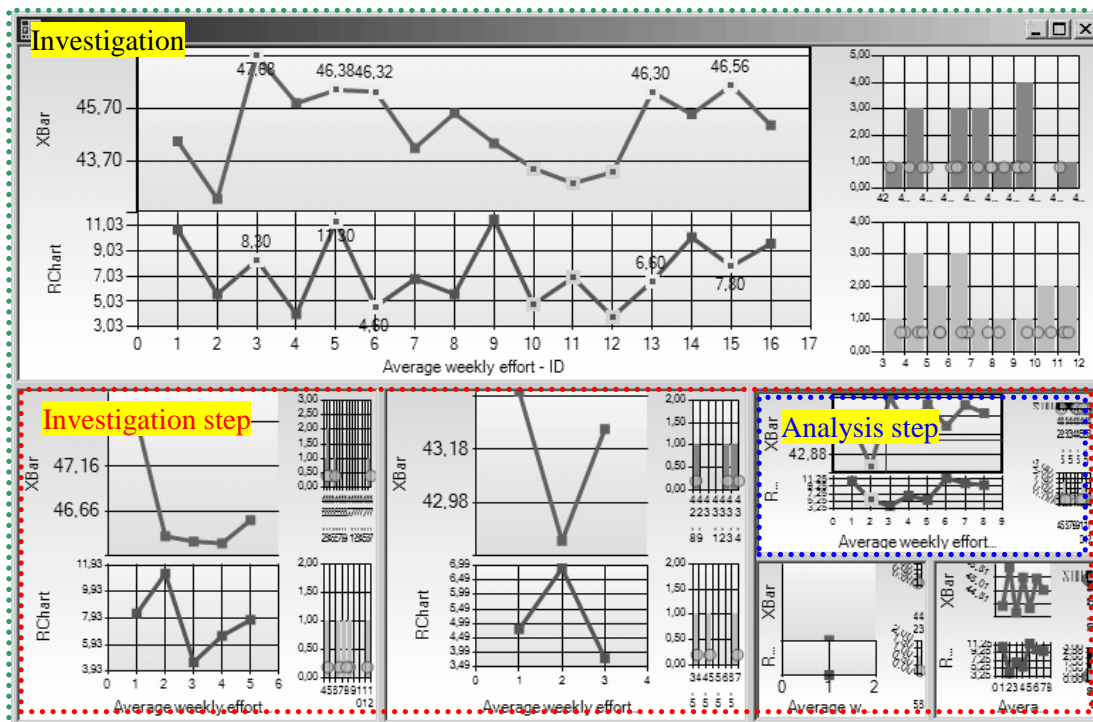


Obrázek 5-3 Diagram tříd logické vrstvy



Obrázek 5-4 Způsob řazení algoritmů (příklad X-Bar/R Chart grafy)

Jednotlivé třídy logické vrstvy navzájem korespondují s pojmy uvedenými dříve. Zároveň jednotlivým třídám také odpovídají prvky z presentační vrstvy. Jejich popis je na Obrázek 5-5.



Obrázek 5-5 Presentační vrstva - popis komponent

6 Ukázka analýzy dat

Tato kapitola přináší ukázkou vzorové analýzy reálných dat. Jsou zde popsány základní fakta o analyzovaném procesu, podrobně rozebrány jednotlivé kroky s odkazem na příslušné video a v závěru jsou shrnuty a diskutovány dosažené výsledky. Poslední podkapitola jen pro úplnost obsahuje stručný popis všech demonstračních videí, která jsou součástí přiloženého DVD.

V úvodu nejdříve popíšeme některá základní fakta. Z pochopitelných důvodů není možné v této práci zmínit všechny dostupné údaje, proto se budeme v některých případech na tuto skutečnost odvolávat. Pro analýzu byla použita reálná data společnosti ANF DATA spol. s r.o., a Siemens Company, která zachycují plánování vývoje jisté aplikace na platformě C#. Cílem analýzy je zjistit, zda je tento proces plánování pod statistickou kontrolou a případně najít možnosti, jak tento proces dále vylepšit.

Plánování vývoje aplikace se děje v tzv. *sprintech*. Každý sprint má délku dvou týdnů a tým na začátku každého sprintu stanoví úkoly, které se pak během této doby snaží splnit. Jednotlivým úkolům přiřadí odhadovaný počet hodin a po splnění daného úkolu se zaznamená čas, který byl opravdu potřeba. Z rozdílu těchto dvou hodnot (odhadnutého času a reálného času) lze stanovit přesnost plánování. Všechny tyto údaje tým zaznamenává do tzv. *backlogu*. Tento dokument obsahuje mimo jiné také další údaje, které budou během analýzy potřeba. Zahrnují například popis jednotlivých úkolů, dodatečné komentáře, záznamy o tom, kdo úkoly provedl a další. S ohledem na zachování anonymity není možné tyto dokumenty zveřejnit. Data z nich byla proto přejata do databáze, která bude sloužit jako výchozí bod analýzy. Popis jednotlivých atributů této analýzy je popsán ve videu, které celou analýzu zachycuje.

6.1.1 Postup analýzy

Celá analýza je rozdělena do několika kroků. Video, které je umístěno na DVD nosiči, zahrnuje všechny kroky analýzy s podrobným komentářem.

V prvním kroku budeme analyzovat celý proces plánování. Pro tento účel použijeme XmmR graf (Median moving range chart, viz. kapitola 3.3.3). Volba tohoto grafu je ovlivněna dvěma faktory. Zprvce analyzujeme jednotlivé záznamy zvlášť, data tedy neseskupujeme. Zadruhé grafy poskytují užší kontrolní meze než XmR grafy, díky čemuž je snáze možné odhalit nestabilitu procesu. Jako zdrojová data k analýze nám poslouží celkový počet odhadnutých hodin na daný sprint, který se skládá jednak z hodin naplánovaných na začátku každého sprintu a také z hodin, které jsou během sprintu dodatečně zařazeny.

Z Obrázek 6-1 plyne, že tento proces je pod statistickou kontrolou. Zároveň lze velmi dobře pozorovat jeho postupnou stabilizaci (spodní graf). Na začátku, kdy byl tento systém plánování zaveden, byly odhady velmi nepřesné. Postupem času a s přibývajícím zkušenostmi se proces plánování stabilizoval. Toto zjištění koresponduje s empirickými zkušenostmi, podle kterých se přibližně po pátém sprintu od zavedení plánování začnou odhady upřesňovat. Během posledních čtyř sprintů lze navíc pozorovat vzestupný trend počtu naplánovaných hodin, který úzce souvisí s přijetím nového člena do týmu.

Z výše uvedené analýzy tedy plyne, že proces je pod statistickou kontrolou. Při bližším pohledu na jednotlivé záznamy (*backlog*) však zjistíme, že některé úkoly byly během sprintu dodatečně zařazeny navíc. Důvody těchto kroků mohou být různé, nejčastěji se však jedná o situaci, kdy tým zvládne splnit všechny úkoly s časovým předstihem. Tyto přidané hodiny mohly pochopitelně celkový proces plánování negativně ovlivnit a v dalším kroku se tedy zaměříme právě na ně.



Obrázek 6-1 První krok analýzy - posouzení celkové stability procesu plánování.

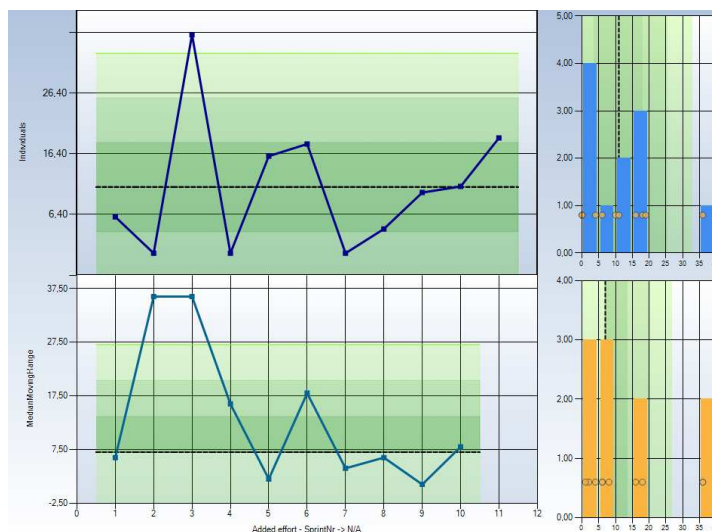
Nyní budeme analyzovat proces dodatečného plánování a přidávání hodin během právě probíhajícího sprintu. Pro analýzu opět použijeme XmmR graf ze stejných důvodů jako výše. Z provedené analýzy (Obrázek 6-2) plyne, že ve třech sprintech (označených čísly 9, 12 a 20) došlo k překročení stanovených kontrolních mezí. Toto je jasným signálem pro zahájení podrobné analýzy. Je nutné se tedy vrátit k původním záznamům a v nich dohledat všechny možné příčiny tohoto nestandardního chování.



Obrázek 6-2 Druhý krok analýzy - dodatečné plánování hodin.

Ze záznamů zjistíme, že ve sprintech číslo 9 a 12 bylo dodatečně naplánováno celkem 137,5 hodin a to na konkrétní typ úkolu, který z pochopitelných důvodů není možné jmenovat. Pro pozdější lepší orientaci jej budeme označovat například jako „specifický typ úkolů“. Zároveň jsme zjistili, že plánování sprintu číslo 20 bylo ovlivněno požadavky zákazníka na odstranění některých chyb (*bug fixing*). Tyto požadavky jsou vzhledem k tomu, že aplikace byla v tomto období ve fázi systémového testování, pochopitelné. Projdeme proto ještě jednou všechny záznamy a pokusíme se v nich identifikovat všechny sprinty, které obsahují podobné vlastnosti – práci na *specifickém typu úkolů* nebo požadavky zákazníka na odstranění chyb. Z této analýzy vyplynulo, že ve sprintech číslo 9 a 12 byly dodatečně naplánovány hodiny navíc vzhledem k práci na *specifickém typu úkolů*. Dále se zjistilo, že sprinty číslo 17, 19 a 20 obsahovaly velké množství dodatečně naplánovaných hodin kvůli

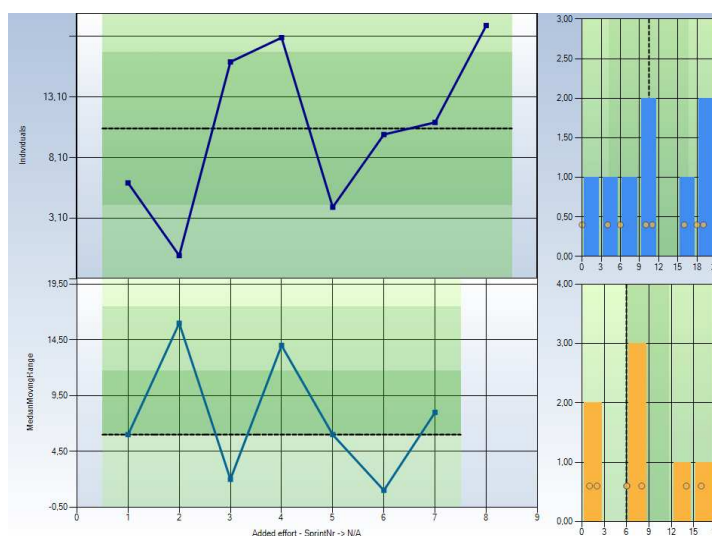
nečekaným požadavkům zákazníka na opravu chyb. Všechny tyto sprinty z právě prováděné analýzy odstraníme, abychom zjistily, zda se zbývající sprinty chovají stabilně. Provedeme tedy další *krok šetření*, který je zachycen na Obrázek 6-3.



Obrázek 6-3 Třetí krok analýzy - po odstranění nevyhovujících sprintů.

Ještě jednou zopakujme, jaké kroky jsme dosud provedli. Analyzovali jsme proces plánování a z grafu (Obrázek 6-1) jsme zjistili, že je pod statistickou kontrolou. Abychom toto tvrzení podpořili nebo vyvrátili, analyzovali jsme počty ke sprintu doplánovaných hodin (Obrázek 6-2). Z této analýzy však vyplynulo několik signálů nestability a proto jsme je odstranili z dalšího kroku analýzy, abychom zjistili, zda se bez nich proces plánování dodatečných hodin chová stabilně.

Nyní k výsledkům, které zachycuje Obrázek 6-3. Z nich je opět vidět, že ve sprintu číslo 7 došlo k překročení horní kontrolní meze. Pokud analyzujeme příčinu tohoto stavu, zjistíme, že tým zvládl splnit všechny naplánované úkoly s velkou časovou rezervou a mohl si proto dovolit na konci sprintu některé úkoly přidat. Tím ovšem porušil proces plánování a ovlivnil tak celkový počet na sprint naplánovaných hodin. Zaměříme tedy svoji pozornost na to, proč tým splnil úkoly ze sprintu číslo 7 tak rychle.



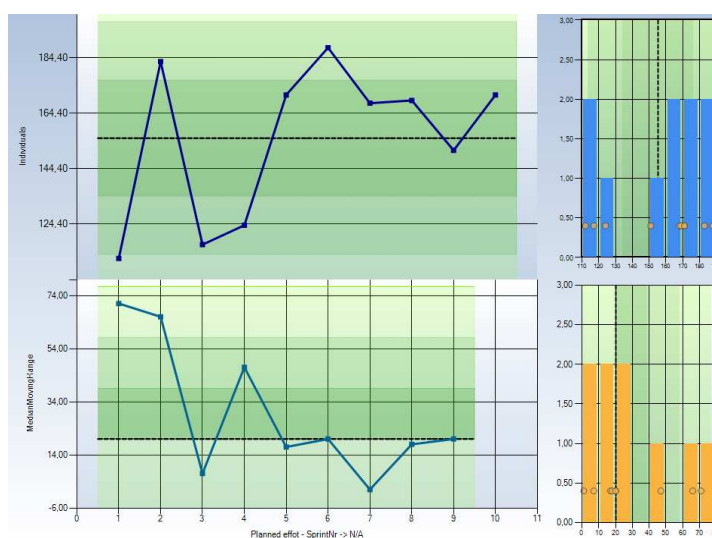
Obrázek 6-4 Čtvrtý krok analýzy - odstraněny sprinty 7, 9 a 13.

Z příslušných záznamů zjistíme, že hlavním důvodem byl fakt, že většina úkolů souvisela s kontrolou návrhu, ověřováním funkcionality a opravou drobných chyb a naopak jen velmi malé úsilí

bylo věnováno samotnému vývoji. Projdeme tedy všechny zbývající sprinty a budeme hledat, zda v některém z nich nedošlo k nahromadění podobných úkolů jako ve sprintu 7. Takto identifikujeme snadno sprinty 7, 9 a 13 a odstraníme je z další iterace. Ta je zachycena na Obrázek 6-4. Z něj vyplývá, že proces plánování dodatečných hodin je ve zbývajících sprintech pod statistickou kontrolou.

Pokud tedy shrneme výsledky dosažené analýzou procesu plánování dodatečných hodin, můžeme konstatovat, že tento proces ovlivňuje celkové plánování a je nutné jej proto stabilizovat. Především je nutné zlepšit plánování *specifických úkolů*, v budoucnu vyhradit více času na případnou opravu chyb a zpřesnit plánování úkolů, které souvisejí s kontrolou návrhu, ověřováním funkcionality a opravou drobných chyb. Během této analýzy také byly odstraněny následující sprinty: 7, 9, 12, 17, 19 a 20. Tyto dílčí závěry vezmeme v potaz a vrátíme se zpět k analýze procesu celkového plánování.

V dalším kroku analýzy odstraníme výše uvedené sprinty z procesu celkového plánování, abychom zjistili, zda plánování v ostatních sprintech je pod statistickou kontrolou. Výsledek tohoto kroku je vidět na Obrázek 6-5. Z něj jasně plyne, že plánování ve zbývajících sprintech je stabilní a lze tedy učinit celkový závěr.



Obrázek 6-5 Pátý krok analýzy - celkové plánování po odstranění sprintů 7, 9, 12, 17, 19 a 20.

6.1.2 Dosažené výsledky

Tato podkapitola shrnuje výsledky ukázkové analýzy reálných dat společnosti ANF Data a poukazuje na možnosti dalšího zlepšování procesu plánování. Zároveň také obsahuje závěry, které byly vyvozeny na základě rozhovoru a vedením této firmy.

Analýzou procesu plánování sprintů ve firmě ANF Data bylo zjištěno, že tento proces je stabilní. Úkoly, které tým na začátku každého sprintu stanoví, společně s úkoly, které během sprintu dodatečně naplánuje, vždy splní v rámci povolených mezí. Analýza dále ukázala postupnou stabilizaci plánování během prvních pěti sprintů, což potvrdilo empirické zkušenosti.

Na základě konzultací k dosaženým výsledkům s vedením firmy byly učiněny následující závěry. Přesnější plánování tzv. „*specifických úkolů*“ bylo z hlediska projektu obtížně zjistitelné a souviselo především s nejasnými požadavky ze strany zákazníka. V tomto případě se jednalo o řešení velmi obtížných úkolů, jejichž výsledky nemohly být dopředu dobře známy.

Dále byla vytvořena odpovídající časová rezerva pro úkoly související s údržbou aplikace a zajištěním jejího bezproblémového chodu. Tímto byl také vytvořen prostor pro případné řešení nečekaných požadavků ze strany zákazníka. Byl také přehodnocen přístup při plánování úkolů souvisejících s ověřováním funkcionality a testováním částí kódu.

Vedení firmy se také rozhodlo analýzu uplatnit na širší skupinu projektů s cílem vylepšit jejich chování a zvýšit efektivitu práce. Cílem tohoto kroku je také prevence a odstraňování systematických chyb z procesů. Jako vedlejší výsledek provedené analýzy byla zjištěna potřeba vytvořit podpůrný systém pro automatizovaný sběr dat. Cílem je zajistit především jejich potřebnou kvalitu a tím i následně zpřesnit výsledky v budoucnu prováděných analýz.

Velmi důležitý závěr byl učiněn v souvislosti s používáním metrik pro analýzu výkonnosti jednotlivých členů týmu. Použití metrik s tímto účelem je naprosto nežádoucí a v souvislosti s projektem působí nevhodně. Metriky proto budou důrazně používány pouze pro analýzu a zlepšování procesů. Tento závěr tedy podporuje teoretická východiska celé práce.

Provedením výše zmíněných úprav lze všechny procesy ve firmě ANF Data do budoucna zefektivnit a zvýšit jejich výkonnost. Po určitém čase lze také postupně zvyšovat jejich vospělost, což je hlavním cílem společnosti.

Metoda statistické analýzy dat a vytvořená aplikace byly celkově shledány jako mimořádně efektivní a poskytující vypovídající způsob pohledu na projekty. Aplikace dokázala ve velmi krátkém čase identifikovat všechna místa, která je možné vylepšit, čímž prokázala svou všestrannost a širokou škálu uplatnění.

6.1.3 Demonstrační videa

Ukázková videa slouží především pro snadné seznámení se s aplikací bez nutnosti ji instalovat nebo spouštět. DVD obsahuje celkem devět krátkých video ukázek doplněných komentáři a jedno delší video (cca 20 minut) zachycující postup při analýze reálných dat. Jednotlivé ukázky na sebe navzájem navazují a postupně tak rozšiřují získané znalosti. Jejich seznam je v Tabulka 6-1 Seznam ukázkových videí. Pro spuštění videí bude v některých případech nutné nainstalovat přehrávač KMPlayer, který je umístěn na DVD.

Název	Délka	Popis
1 Zaklady prace.avi	4:23	Rozdělení aplikace, základy práce s aplikací, definice analýz
2 Zjednodusene zadavani analyz.avi	2:21	Popis zjednodušeného zadávání analýz – pomocník obsahující předem navolené modelové situace
3 Synchronizace grafu.avi	2:06	Ukázka synchronizace výběru dat v rámci analýzy
4 Kontrolni meze.avi	2:14	Popis práce s kontrolními mezemi grafu, volba zobrazení,
5 Prirazovani do skupin.avi	4:20	Přiřazování dat do skupin, definice skupin, automatické přiřazení dat,
6 Zobrazeni dat analyzy, vyber hodnot.avi	2:24	Zobrazování dat k analýze, výběr dat a synchronizace s grafem, řazení dat
7 Kroky analyzy.avi	3:36	Vytváření kroků analýzy, identifikace grafů podle barvy skupiny, zvětšení grafu
8 Docasne kontrolni meze.avi	1:06	Zobrazení dočasných kontrolních mezí
9 Pokrocila prace s grafy.avi	1:16	Pokročilá práce s grafy, přibližování grafů
Analyza realnych dat.avi	19:17	Vzorová ukázka analýzy reálných dat
43:09		

Tabulka 6-1 Seznam ukázkových videí.

7 Závěr

Tato práce přinesla přehled všech nejdůležitějších poznatků z oblasti managementu procesů se zaměřením na jejich zlepšování a zefektivňování.

V jejím úvodu byly vymezeny základní pojmy a stručně nastíněna problematika zlepšování softwarových procesů. Dále byla věnována pozornost problémům, které mohou nastat během zavádění procesů do praxe. Zmíněn byl především negativní vliv lidského faktorku, který hraje zásadní roli prakticky ve všech fázích životního cyklu procesu. Důraz zde byl kladen také na správnou a úplnou definici metrik, které následně slouží jako zdroj dat pro analýzu.

Hlavní část práce byla věnována popisu metody statistického přístupu k analýze chování procesů. Zdůrazněn byl především význam stability procesu vzhledem k jeho předvídatelnosti a byly zde popsány dvě hlavní příčiny tohoto chování. Práce také obsahuje popis všech typů grafů.

Všechny výše uvedené teoretické znalosti pak byly uplatněny při návrhu a realizaci aplikace, která se soustřeďuje právě na analýzu dat pomocí grafů. Koncepte aplikace zahrnuje nejenom tyto teoretické poznatky, ale snaží se zejména přinést nové a inovativní přístupy, v čemž spočívá hlavní přínos práce. Mezi ty nejdůležitější patří generický návrh a možnost analyzovat prakticky libovolný proces, což je navíc podpořeno možností připojit aplikaci prakticky k libovolné databázi. Do uživatelského rozhraní aplikace bylo dále zabudováno několik prvků, které pomáhají proces analýzy usnadnit a většinu kroků automatizovat. Jmenujme například pomocníka pro definici analýz, komponentu, která ověřuje jejich platnost, automatické funkce rozdělování bodů do skupin podle uživatelem zadaných kritérií a v neposlední řadě také plně automatické a přehledné vytváření nových kroků analýz.

Aplikace byla následně otestována na reálných datech a výsledky, které přinesla, ověřily její funkčnost a zároveň potvrdily použitelnost metody statistické analýzy při zlepšování softwarových procesů. Všechny závěry provedených analýz byly následně implementovány do praxe. Tento krok tak pomohl vylepšit analyzovaný proces, což byl jistě další přínos celé práce.

Z pohledu dalšího vývoje je možné doplnit práci o kapitoly, které se budou podrobněji věnovat problematice vspělosti procesu (*process capability*) a zaměřit se zde především na různé způsoby vylepšování procesů.

Literatura

- [FAJ] FAJMON, B., RUZICKOVA, I. Matematika 3. Skriptum matematiky, Fakulta elektrotechniky a komunikačních technologií VUT v Brně, 254 stran.
- [FLO99] FLORAC, William A., CARLETON, Anita D. *Measuring the software process: Statistical process control for software process improvement*. [s.l.] : [s.n.], 1999. 250 s.
- [HAR60] HARTER, H. L. "Tables of Range and Studentized Range," *Annals of Mathematical Statistics* 31 (1960).
- [CHA87] CHAMP, W., WOODALL, H. *Exact results for Shewhart Control Charts with Supplementary Runs Rules*. *Technometrics* 29, no. 4 (1987): 393-399.
- [IBM09] IBM. *Rational Unified Process* [online]. 2009 [cit. 2009-01-05]. Dokument dostupný na URL: <http://www-01.ibm.com/software/awdtools/rup/>.
- [WEE58] Western Electronic Co., Inc. *Statistical Quality Control Handbook*. Indianapolis: AT&T Technologies, 1958.
- [WHE92] WHEELER, Donald J., CHAMBERS, David S. *Understanding Statistical Process Control*. Second Edition. Knoxville, Tennessee: SPC Press, 1992.
- [WHE95] WHEELER, Donald J. *Advanced Topics in Statistical Process Control*. Knoxville, Tennessee: SPC Press, 1995.
- [UP] *Unified process*. Wikipedia [online], 2009 [cit. 2009-05-01]. Dokument dostupný na URL: http://en.wikipedia.org/wiki/Unified_Process.
- [VAC09] VACULNÝ, J. *Agilní metodologie programování*. MU [online] 2009. Dokument dostupný na URL: <http://www.fi.muni.cz/usr/jkucera/pv109/2003/xvaculny.htm>
- [VER09] VERNER, J. *Metriky procesů vývoje softwaru*. Brno, 2009. Semestrální projekt, FIT VUT v Brně.
- [ZEN06] ZENDULKA, J., BARTÍK, V., LUKÁŠ, R., RUDOLFOVÁ, I. *Získávání znalostí z databází*. Studijní opora Fakulta informačních technologií VUT v Brně (160 stran), 2006.

Seznam příloh

Příloha 1. DVD Metriky procesů vývoje softwaru

Příloha obsahuje

- Zdrojové kódy aplikace pro analýzu dat – složka *./Source*
- Ukázková videa, která zahrnují celkový popis aplikace – složka *./Tutorials*
 1. Zaklady prace.avi 4:23 Rozdělení aplikace, základy práce s aplikací, definice analýz
 2. Zjednodusene zadavani analyz.avi 2:21 Popis zjednodušeného zadávání analýz – pomocník obsahující předem navolené modelové situace
 3. Synchronizace grafu.avi 2:06 Ukázka synchronizace výběru dat v rámci analýzy
 4. Kontrolni meze.avi 2:14 Popis práce s kontrolními mezemi grafu, volba zobrazení
 5. Prirazovani do skupin.avi 4:20 Přiřazování dat do skupin, definice skupin, automatické přiřazení dat
 6. Zobrazeni dat analyzy, vyber hodnot.avi 2:24 Zobrazování dat k analýze, výběr dat a synchronizace s grafem, řazení dat
 7. Kroky analyzy.avi 3:36 Vytváření kroků analýzy, identifikace grafů podle barvy skupiny, zvětšení grafu
 8. Docasne kontrolni meze.avi 1:06 Zobrazení dočasných kontrolních mezí
 9. Pokrocila prace s grafy.avi 1:16 Pokročilá práce s grafy, přibližování grafů
Přehrávač KMPlayer The_KMPlayer_1434.exe
- Video, které zachycuje analýzu reálných dat – složka *./Demo*
 - Analyza realnych dat.avi 19:17 Vzorová ukázka analýzy reálných dat
 - Přehrávač KMPlayer The_KMPlayer_1434.exe
- Knihovny potřebné pro spuštění aplikace – složka *./Libraries*
- Text této práce – složka *./Text*