

# VÝPOČETNÍ MODEL INTELIGENTNÍHO VÝBĚRU TESTOVÝCH OTÁZEK

## THE COMPUTING MODEL FOR INTELLIGENT SELECTION OF THE TESTING QUESTIONS

PETR VOBORNÍK

Fakulta informatiky a managementu / Univerzita Hradec Králové

---

### Abstrakt

Výběr otázek pro sestavování periodicky zadávaných opakovacích testů může probíhat zcela náhodně, ale také může mít důmyslně propracovaný postup, který zvýší jejich použitelnost. V tomto článku bude předveden právě jeden z takovýchto modelů, který se pokusí vhodným způsobem zkombinovat všechny relevantní složky a optimalizovat tak výběr otázek, aby efektivita testů byla pro daný účel co nejvyšší. Zároveň může být tento výběrový model inspirací i v jiných vědních oborech.

### Klíčová slova

Testy, otázky, skupiny, výběr, pravděpodobnost, křivka zapomínání, ruletová selekce.

### Abstract

Selection of the questions for the compilation of periodically repeating tests may be completely random. Also it may be a sophisticated process which will increase their usability. In this article will be introduced one of these models. These model attempts to appropriately combined all relevant components for optimize the selection of questions to test and raise the effectiveness for purpose. This selection model can also be inspiration in other disciplines too.

### Key words

Tests, questions, groups, selection, probability, forgetting curve, roulette wheel selection.

**Školitel:** prof. RNDr. Eva Milková, Ph.D.

## Úvod

V projektu Univerzální testovací prostředí, představeném v článku [1], je jednou z jeho stěžejních součástí schopnost sestavovat vždy originální test s náhodně vybranými otázkami. Počet otázek, jež mají tvořit jednotlivé testy, obvykle bývá nižší, než je počet otázek, které jsou pro daný test k dispozici. Z takovéto databáze se pak náhodně vybírají otázky pro každý spuštěný test. Může jít o výběr např. 5 z 6, ale také z 500. Tento výběr ovšem nemusí být jen zcela náhodný, naopak v některých případech je žádoucí umožnit náhodnost výběru otázek ovlivnit tak, aby lépe plnila účel, pro který je test používán. Mezi účely vyžadující různé typy míchání při tom patří např. testování výukové, drilovací, opakovací či hodnocené závěrečné.

Zcela náhodné míchání je třeba při závěrečném hodnoceném testování. U výukového testování by zase měly být upřednostňovány nové, zkoušeným dosud neřešené otázky. U testování drilovacího je naopak třeba, aby se již probrané otázky opakovaly ve specifických intervalech odpovídajících individuálním křivkám zapomínání jednotlivých zkoušených. Opakovací testování by pak při výběru mělo vhodným způsobem brát v potaz míru správnosti předchozích řešení každé otázky a zároveň dobu od jejího zodpovězení. A právě tímto posledním případem se budeme v tomto článku zabývat.

Opakovací testy kombinují didaktické testy ověřující a průběžné [2 str. 16]. Měly by rychle a objektivně prověřit znalost zkoušených (studentů) [3 str. 23]. Dovolují-li to podmínky, měly by být zadávány v pravidelných periodách, optimálně při každé hodině či cvičení z daného předmětu (tzv. „pětiminutovky“). Jejich obsahem by měla být předně látka probraná v předchozím cvičení, ovšem alespoň částečné zastoupení by mělo patřit i látce předešlé. To platí samozřejmě jen v některých předmětech, kde slouží tyto testy nejen k prověření aktuálního stavu znalostí studentů, ale i jako motivace k průběžnému studiu a přípravě na výuku.

## 1 Inteligentní výběr testových otázek

Pro optimální výběr otázek do periodicky zadávaného opakovacího testu je zapotřebí zohlednit několik faktorů. Ty by sice měly ovlivňovat pravděpodobnost výběru, avšak nikdy z něj žádnou otázku zcela nevyřazovat. Mezi tyto faktory patří počet řešení otázky v minulosti, doba uplynulá od těchto řešení a dosažený výsledek. Jejich vhodné zkombinování pro optimální výběr otázek bude předmětem této části.

### 1.1 Složka zapomínání

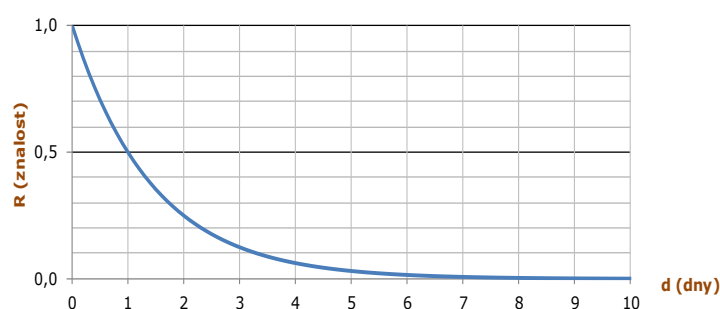
Složka zapomínání by měla zohledňovat dobu, jež uplynula od předchozího řešení dané otázky. Předpokladem je, aby nedávno zadaná otázka měla vyšší pravděpodobnost, že do nového výběru nebude zařazena, než ta, která byla řešena dříve.

Pro stanovení tohoto časového vlivu se přímo nabízela tzv. křivka zapomínání, jejíž tvar zhruba odpovídá předpokladům využití a zároveň má i své vědecké opodstatnění z oblasti psychologie. Autorem této teorie je Hermann Ebbinghaus [4 str. 94-96], který ve své práci [5], uvádí, že nejrychleji zapomínání probíhá v prvních hodinách od naučení se určitého faktu, přičemž do druhého dne jde až o 80% znalostí [6 str. 201].

Matematického modelu křivky zapomínání je zároveň využíváno ve výukových softwarech, kterými se uživatelé učí drilovací metodou, tj. opakováním faktů s vhodně zvolenou prodlevou mezi jednotlivými opakováními (tzv. „spaced repetition“, viz [7 str. 6-8]).

Z této teorie vychází i metoda Re-wise používaná např. firmou LANGMaster. Tato křivka předpokládá, že nově naučená fakta jsou zapomínána geometrickou řadou, konkrétně polovina z aktuálně pamatovaných faktů denně. Pomineme-li tedy efekt opakování a individuálnost jednotlivých osob a jejich vztahu k různým faktům, pak tuto křivku zapomínání vyjadřuje vzorec (1) a znázorňuje Graf 1. [8 str. 20-21]

$$R = \frac{1}{2^d} \quad (1)$$

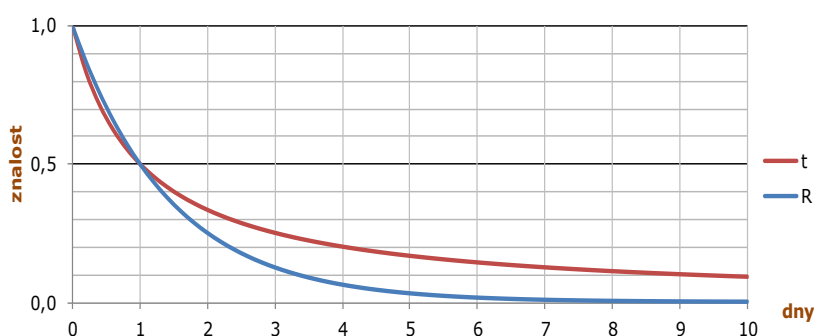


Graf 1: Křivka zapomínání dle metody Re-wise

Tato verze křivky zapomínání ovšem velmi rychle klesá a již po čtyřech dnech odhaduje zapamatované znalosti pouze na 6,25%, po 7 dnech pak na pouhých 0,78%. Při takto rychlém vytěsnění časového efektu je tato křivka pro řešený model výběru otázek v podstatě nepoužitelná. Lze sice např. exponent jmenovatele **d**, vyjadřující počet dnů od naučení se sledovaného faktu, vydělit určitou hodnotou a zmírnit tak počáteční strmý pokles, avšak i po této úpravě je křivka stále příliš strmá pro dosažení kýženého efektu.

Hodnocený opakovací test totiž nemá za cíl zkoušené učit, ale naopak periodicky ověřovat jejich znalosti. Proto navržený matematický model ovlivňující výběr otázek nevychází z křivek zapomínání [9], ale snaží se více reflektovat potřebu obměny otázek a pokrytí celého spektra testovaného učiva. Místo geometrické řady byla tedy časová proměnná přesunuta pouze do jmenovatele (2), což pro složku zapomínání **t** představuje zřetelně plošší křivku (pomaleji klesající) než v předchozím případě (viz Graf 2).

$$t = \frac{1}{d+1} \quad (2)$$

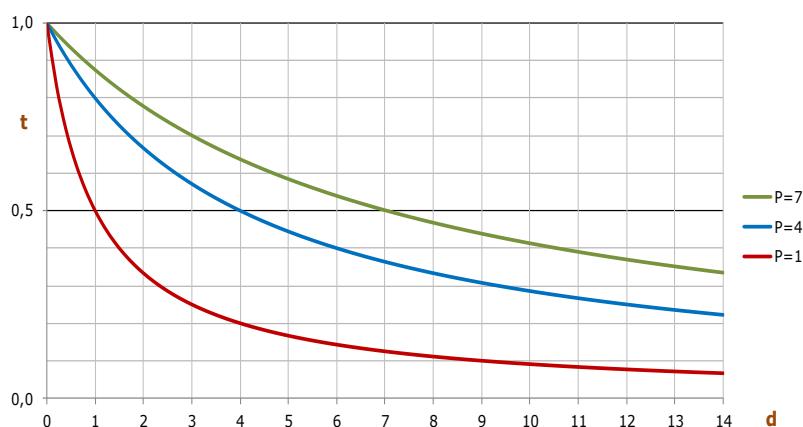


Graf 2: Porovnání křivky zapomínání **R** (1) a navrhované křivky složky zapomínání **t** (2)

Složka zapomínání označená  $t$  tedy vyjadřuje míru zapomínání v čase ve zprava uzavřeném intervalu od 0 (žádná znalost) do 1 (plná znalost). Vzorec (2) byl následně ještě doplněn o parametr  $p$ , který umožňuje rychlost klesání křivky přizpůsobit periodě opakování testu (3).

$$t = \frac{p}{d+p} \quad (3)$$

Parametr  $p$  je tedy průměrná perioda mezi jednotlivými testy vyjádřena ve dnech (např. 7 pro „jednu týdně“). Doslovný význam parametru  $p$  je: „Po kolika dnech má mít minule zcela správně zodpovězená otázka 50% pravděpodobnost zařazení do výběru oproti otázkám novým.“ Graf 3 ukazuje průběhy jednotlivých křivek  $t$  pro tři různé hodnoty parametru  $p$ .



Graf 3: Průběh křivek  $t$  pro různé hodnoty parametru  $p$  dle vzorce (3)

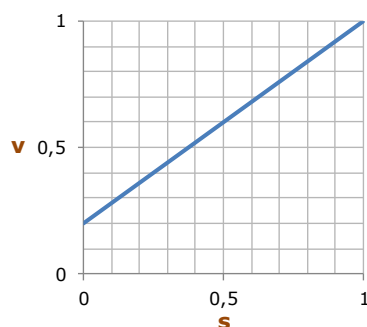
## 1.2 Znalostní složka

Dalším faktorem, který je nezbytné při výběru otázek zohlednit je jejich předchozí výsledek. Na rozdíl od učení se prostým faktům, kde je výsledek obvykle dvoustavový (věděl – nevěděl), je zde výsledkem řešení jednotlivých otázek reálná hodnota z fuzzy stupnice v intervalu  $\langle 0, 1 \rangle$ .

K výslednému skóre je však zapotřebí ještě připočítat „znalost“, kterou dotyčný získal díky řešení této otázky. Po testu totiž může následovat podrobný rozbor řešení jednotlivých otázek, v němž se student vždy dozví, co udělal chybně a jak a proč to mělo být správně. Díky tomu znalost zkušeneho dané otázky, ač ji třeba zodpověděl zcela špatně (na 0%) vzroste nad nenulovou úroveň. Pro vyjádření tohoto „poučení se z vlastních chyb“ je do výpočtu zařazen parametr  $n$ , jež vyjadřuje tuto hodnotu.

$$v = \frac{s+n}{1+n} \quad (4)$$

Vzorec (4) tedy určuje míru předpokládané znalosti tématu otázky  $v$  po jejím zodpovězení a zobrazení rozboru (zpětné vazby). Proměnná  $s$  je zde skóre otázky získané za danou otázku v čase před  $d$  dny. Hodnota  $v$  je tedy navýšena o část  $n$  tak, že výsledek je opět v rozsahu  $\langle 0, 1 \rangle$ . Graf 4 ukazuje vztah mezi těmito dvěma hodnotami ( $s$  a  $v$ ) pro hodnotu parametru  $n = 0,25$ .



Graf 4: Vztah hodnot  $v$  a  $s$  pro  $n = 0,25$

### 1.3 Váhy pro výběr

Předchozí dvě vypočtené hodnoty, tj. složka zapomínání  $t$  (3) a znalostní složka  $v$  (4) jsou následně zkombinovány jejich součinem do hodnoty  $z$  (5).

$$z = t \cdot v \quad (5)$$

Výsledkem  $z$  je aktuální předpokládaná znalost tématu otázky, vzhledem k jejímu dřívějšímu řešení. V případě, že otázka byla řešena vícekrát ( $c$ -krát), lze obecně tento vzorec pro  $i$ -té řešení ( $i \in \{1, 2, \dots, c\}$ ) rozepsat následovně (6).

$$z_i = \frac{p}{d_i + p} \cdot \frac{s_i + n}{1 + n} \quad (6)$$

Výpočet celkové aktuální znalost  $Z$  z více různých  $z_i$ , tj. vícenásobného řešení téže otázky, se určí jako sloučení pravděpodobností jednotlivých jevů  $z_i$ . Tento postup lze vyjádřit pomocí rekurze (7), přičemž konečnou pevnou hodnotou je  $f(0) = 0$ , tzn. v případě, že uživatel tuto otázku ještě nikdy neřešil ( $c = 0$ ), bude výsledkem funkce a tedy i  $Z$  rovno 0.

$$f(x) = z_x + f(x - 1) - z_x \cdot f(x - 1) = z_x + (1 - z_x) \cdot f(x - 1), \quad f(0) = 0 \quad (7)$$

Celková aktuální znalost otázky  $Z$ , přičemž tuto otázku uživatel řešil  $c$ -krát, je tedy určena výše popsanou funkcí (7), kde jejím parametrem bude  $c$  (8).

$$Z = f(c) \quad (8)$$

Váha pro zařazení otázky do výběru, kterou označíme  $P$ , je pak doplňkovou hodnotou pravděpodobné znalosti otázky  $Z$  (9).

$$P = 1 - Z \quad (9)$$

Hodnota  $P$  tedy vyjadřuje váhu pro zařazení otázky do výběru, přičemž váhu 1 mají otázky dosud neřešené, ty ostatní ji mají nižší. Vzhledem k tomu, že jeden zkoušený nebude psát nikdy více testů v témže čase, bude složka zapomínání  $t$  vždy menší než 1, takže  $P$  nebude nikdy nulové ( $t < 1 \Rightarrow Z < 1 \Rightarrow P > 0$ ). Díky tomu, nebude žádná otázka z výběru nikdy zcela vyřazena.

Tab. 1 ukazuje postup výpočtu hodnoty  $P$  pro otázku, jež byla zkoušeným v minulosti již 4x řešena v uvedených datech a s uvedenými výsledky  $s$ .

Datum	d	t	s	v	z	Z (kumulativně)
25.01.2012 10:30	8,063	0,4819	<b>0,8</b>	0,84	0,40482	0,40482
19.01.2012 12:00	14,000	0,3488	<b>0,5</b>	0,60	0,20930	0,52939
12.01.2012 11:05	21,038	0,2628	<b>0,3</b>	0,44	0,11563	0,58381
25.09.2011 09:00	130,125	0,0545	<b>0,4</b>	0,52	0,02834	<b>0,59560</b>
Pro aktuální datum a čas 2.2.2012 12:00; $k = 7,5$ ; $n = 0,25$						<b>P = 0,40440</b>

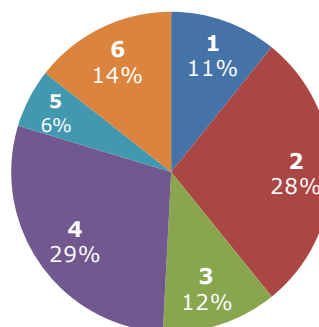
Tab. 1: Ukázka výpočtu hodnoty **P** (9) pro jednu otázku

## 1.4 Výběr otázek podle vah

Sestavení testu probíhá náhodným výběrem ze všech dostupných otázek, přičemž pravděpodobnost výběru kterékoli z nich je přímo odvozena z ohodnocení **P**. Tým algoritmus výběru se používá v genetických algoritmech pro tzv. selekci individuí a je znám jako ruletový mechanismus („roulette wheel selection“ [10 str. 28-29]). Při tomto postupu je každé otázce přiřazena kruhová výseč, jejíž plocha je přímo úměrná ohodnocení otázky **P** [11 str. 22-23]. Poté se pomyslné kolo rulety roztočí a je vybrána ta otázka, na jejímž úseku se kulička rulety zastaví [12 str. 6]. Zvolená metoda selekce je pro tuto úlohu vhodná, jelikož zde nehrozí limitace jejími nedostatky, jako je např. požadavek na nezápornost řešení [13 str. 23].

Otázka	P	Podíl $\frac{P_i}{\sum P}$	Kumulativně	
			P	Podíl
1	0,375	0,108	0,375	0,108
2	0,984	0,284	1,359	0,392
3	0,404	0,117	1,763	0,508
4	1,000	0,288	2,763	0,797
5	0,206	0,059	2,969	0,856
6	0,500	0,144	3,469	1,000
$\Sigma$	<b>3,469</b>	<b>1,000</b>		

Tab. 2: Ukázka výpočtu velikosti výseče rulety



Graf 5: Ukázka rulety jako koláčového grafu

Výše popsané „točení rulety“ lze programově realizovat mnoha způsoby. Jedním z nich je skutečně převést ohodnocení **P** každé otázky na jednotnou míru, a to vydělením každé z nich jejich celkovým součtem (viz 3. sloupec „Podíl“ v Tab. 2). Výsledek pak v součtu dává hodnotu 1, přičemž generátory pseudonáhodných hodnot nejčastěji generují desetinnou hodnotu právě v rozsahu  $(0, 1)$  z rovnoměrného rozdělení. Kumulace těchto podílů (viz poslední sloupec v Tab. 2) pak vždy udává dolní hranici intervalu, do něhož spadá-li vygenerované náhodné číslo, označuje jím otázku zařazenou do výběru (např. náhodná hodnota 0,637 by dle Tab. 2 označovala otázku č. 4). Na rozdíl od výběru individuí pro křížení v genetických algoritmech se ovšem jedna otázka v tomtéž testu nesmí vyskytovat více než jednou. Proto je pro výběr další otázky třeba celý postup znovu opakovat s vynecháním již zvolené otázky.

Méně náročnou variantou je vynechat přepočtení ohodnocení **P** na jejich podíl a použít rovnou kumulace těchto hodnot (viz předposlední sloupec v Tab. 2). Pseudonáhodně vygenerovanou hodnotu v rozsahu  $(0, 1)$  z rovnoměrného rozdělení pak stačí vynásobit  $\Sigma P$

a výsledek již přímo označuje interval vymezený těmito kumulativními hodnotami. Tím se, mimo další režie při výpočtu, ušetří počet  $\frac{q(2Q-q+1)}{2}$  (kde **Q** je celkový počet otázek v databázi a **q** je počet otázek vybíraných pro test) výpočetně nejnáročnějších operací dělení [14 str. 23], které jsou nahrazeny pouze **q** počtem operací násobení. (Např.  $0,637 \cdot 3,469 \doteq 2,21$ , čili i tímto postupem by byla vybrána otázka č. 3.)

## 2 Skupiny otázek

Výše popsaný postup se týká otázek, jež mají při prvním testu zkoušeného všechny stejnou pravděpodobnost na výběr (**P** = 1). V testech však bývá žádoucí, aby v nich určité skupiny otázek byly zastoupeny předem stanoveným počtem (např. 4 otázky z tématu minulé hodiny a 1 z čehokoli předešlého). Za tímto účelem některé testovací aplikace podporují dělení otázek do skupin, u kterých se zároveň definuje počet otázek, určující maximální zastoupení skupiny v testu.

V prezentovaném univerzálním testovacím prostředí je členění těchto skupin hierarchické, přičemž hloubka úrovní této hierarchie není nijak omezena (viz Obr. 1). Výběr otázek do testu se pak provádí pomocí algoritmu prohledávání do hloubky, kdy je z každé již dále nedělené skupiny vybrán nastavený počet otázek pomocí zde popsaného postupu. Tato skupina je pak zrušena a z ní vybrané otázky přesunuty do vyšší složky. Tímto způsobem se postupuje až do nejvyšší kořenové složky, z níž jsou zvoleny otázky pro samotný test.

Název	Úroveň	Váha	Vždy	Otáček	Časové omezení
Cykly - výstupy	5		<input type="checkbox"/>	1	
Výstupní hodnoty - dokud	5	5	<input type="checkbox"/>		
Výstupní hodnoty - pro	5	5	<input checked="" type="checkbox"/>		00:02:00
Cykly - doplňovačky	5		<input type="checkbox"/>	1	
Po jestliže			<input type="checkbox"/>	3	
If - Výstupy	4		<input type="checkbox"/>	1	
Vykonané řádky	4		<input type="checkbox"/>	1	
Multi select výrazy	4		<input checked="" type="checkbox"/>	1	
Bool výrazy	4	6	<input type="checkbox"/>		
Validita příkazů	4	6	<input checked="" type="checkbox"/>		
Přiřazování	1,4		<input type="checkbox"/>	1	
Úplné větvení	4	4	<input type="checkbox"/>		
Principy algoritmů	1	2	<input type="checkbox"/>		00:01:30
Posun a prohození	3		<input type="checkbox"/>	1	
Prohození proměnných	3	2	<input type="checkbox"/>		
Count=30		Avg=3,17			

Obr. 1: Ukázka hierarchického členění otázek do skupin v administračním rozhraní aplikace

## Závěr

Výběr otázek pro sestavování periodicky zadávaných opakovacích testů může probíhat zcela náhodně, ale také může mít důmyslně propracovaný postup, který zvýší jejich použitelnost. Jeden z takovýchto postupů byl představen i v tomto článku. I když se jedná o model použitelný pouze pro specifický druh testu a pouze pro některé učební předměty, může být inspirací nejen pro implementace budoucích testovacích aplikací, ale i pro další rozvoj teorií výběrových funkcí, ať již v tomto či zcela jiném odvětví.

## Literatura

- [1] VOBORNÍK, P.: *Univerzální testovací prostředí*. In: Sborník příspěvků z konference eLearning 2011, Hradec Králové : Gaudeamus UHK, 2011, str. 80-85, ISBN 978-80-7435-153-2. [Online] 10.11.2011 [citace: 3.1.2012] [http://www.researchgate.net/profile/Petr\\_Vobornik/publications/](http://www.researchgate.net/profile/Petr_Vobornik/publications/).
- [2] CHRÁSKA, M.: *Didaktické testy*. Brno : Paido, 1999, 91 s., ISBN 80-85931-68-0.
- [3] SCHINDLER, R.: *Rukověť autora testových úloh*. Praha : Centrum pro zjišťování výsledků vzdělávání, 2006, 86 s., ISBN 80-239-7111-5. [Online] 3.5.2006 [citace: 2.1.2012] [http://www.ceremat.cz/download.php?id\\_document=1404034187&at=1](http://www.ceremat.cz/download.php?id_document=1404034187&at=1).
- [4] PLHÁKOVÁ, A.: *Dějiny psychologie*. Praha : Grada Publishing, 2006, 328 s., ISBN 978-80-247-0871-3.
- [5] EBBINGHAUS, H.: *Memory: A Contribution to Experimental Psychology*. [překl.] H. A. Ruger a C. E. Bussenius, New York : Teachers College, Columbia University, 1885.
- [6] SCANDURA, J., M., BRAINERD, CH., J.: *Structural/process models of complex human behavior*. Banff, Alberta, Kanada : Springer, 1978, 624 s., ISBN 90-286-0578-9.
- [7] VOBORNÍK, P.: *Počítačové testovací systémy*. In: Sborník příspěvků z konference Alternativní metody výuky 2011, Praha : Univerzita Karlova, 28. 4. 2011, ISBN 978-80-7435-104-4. [Online] 28.4.2011 [citace: 3.1.2012] <http://everest.natur.cuni.cz/konference/2011/prispevky.php#p34>.
- [8] STÁRA, L.: *Software pro výuku cizího jazyka s využitím matematických modelů*. Praha : ČVUT, 2010, Bakalářská práce. [Online] 8.1.2010 [citace: 2.1.2012] [https://dip.felk.cvut.cz/browse/pdfcache/staralad\\_2010bach.pdf](https://dip.felk.cvut.cz/browse/pdfcache/staralad_2010bach.pdf).
- [9] MOHAMAD, J., MAURICE, B.: *A comparative study of learning curves with forgetting*. Applied Mathematical Modelling, New York : Elsevier Science, August 1997, volume 21, issue 8, str. 523-531, ISSN 0307-904X.
- [10] MAULIK, U., BANDYOPADHYAY, S., MUKHOPADHYAY, A.: *Multiobjective Genetic Algorithms for Clustering (Applications in Data Mining and Bioinformatics)*. Heidelberg : Springer, 2011, ISBN 978-3-642-16614-3.
- [11] HYNEK, J.: *Genetické algoritmy a genetické programování*. Praha : Grada Publishing, 2008, ISBN 978-80-247-2695-3.
- [12] JEDLIČKA, M.: *Využití genetických algoritmů pro optimalizaci výběru sanačních technologií*. Brno : Masarykova univerzita, 2006, diplomová práce. [Online] 1.1.2006 [citace: 3.1.2012] [http://is.muni.cz/th/39464/fi\\_m/xjedlic4-dp.pdf](http://is.muni.cz/th/39464/fi_m/xjedlic4-dp.pdf).
- [13] KRPATA, P.: *Separace signálů vibrací točivých strojů s využitím genetických algoritmů*. Praha : ČVUT, 2010, dipl. práce. [Online] 2010 [cit.: 3.1.2012] [http://measure.feld.cvut.cz/cs/system/files/files/cs/vyuka/zaverecne\\_prace/DP\\_2010\\_Krpata\\_Pavel\\_locked.pdf](http://measure.feld.cvut.cz/cs/system/files/files/cs/vyuka/zaverecne_prace/DP_2010_Krpata_Pavel_locked.pdf).
- [14] NARANBAATAR, N.: *Vybrané problémy výpočtů na PC s pohyblivou řádovou čárkou*. Pardubice : Univerzita Pardubice, 2010, bakalářská práce. [Online] 2010 [citace: 4.1.2012] [http://dspace.upce.cz/bitstream/10195/37488/1/NaranbaatarN\\_VybraneProblemy\\_JR\\_2010.pdf](http://dspace.upce.cz/bitstream/10195/37488/1/NaranbaatarN_VybraneProblemy_JR_2010.pdf).