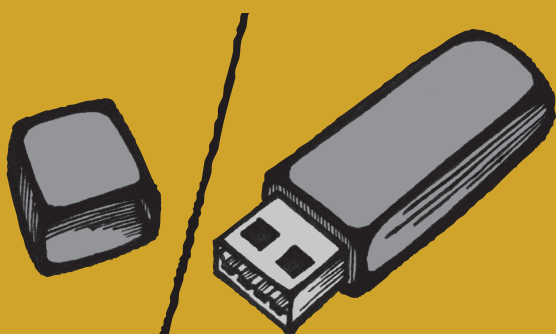


FAKULTA  
APLIKOVANÝCH VĚD  
ZÁPADOČESKÉ  
UNIVERZITY  
V PLZNI

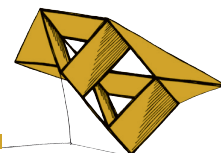
**SBORNÍK**

# STUDENTSKÁ VĚDECKÁ KONFERENCE



BAKALÁŘSKÉ  
STUDIJNÍ PROGRAMY

**30.5.2019**



# Bakalářské studijní programy

## **Sborník rozšířených abstraktů**

Název: SVK FAV 2019 – bakalářské studijní programy

Editor: Jan Rendl

Ilustrace na obálce: Jana Trávníčková

Vydavatel: Západočeská univerzita v Plzni, Univerzitní 8, 306 14 Plzeň

Datum vydání: květen 2019

ISBN 978-80-261-0866-5

Tato publikace neprošla redakční ani jazykovou úpravou.

## Hlavní sponzoři konference



Plzeňský Prazdroj

IEEE  
young  
professionals



UNICORN



REX  
CONTROLS

## Další sponzoři konference



EURO SOFTWARE

A Member of the GK Software Group



aimtec



ŠKODA JS a.s.

CCA<sup>®</sup>  
G R O U P



GEOREAL

PALAXO  
THE WAY FORWARD

PSi

# Obsah

*Sekce – Fyzika, matematika, geomatika, mechanika*

<b>Modální analýza kostry satelitu Pilsen Cube</b> Linhart Jakub	7
---	---

*Sekce – Informatika, kybernetika*

<b>Metody strojové klasifikace pro výběr optimálního estimátoru křivosti</b> Hácha Filip	10
---	----

<b>Detekce tématu dokumentu</b> Kandyba Jan	12
--	----

<b>Elektronický zápisník pro diabetiky</b> Koldovská Aneta	14
---	----

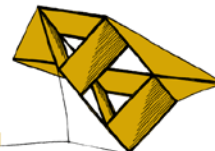
<b>Full-textové vyhledávání s podporou porozumění textu dotazu</b> Mištera Adam	16
--	----

<b>Systém pro rozesílání cvičných phishingových zpráv</b> Šebela Martin	18
--	----

<b>Autonomní navigace kolového robota</b> Švec Josef	20
---	----

## *Sekce*

Fyzika, matematika, geomatika, mechanika



## Modální analýza kostry satelitu Pilsen CUBE

Jakub Linhart<sup>1</sup>

### 1 Úvod

Pilsen CUBE je projekt studentského satelitu typu CubeSat, jehož cílem je nasměrovat mladé lidi k vesmírnému programu. Studenti ve spolupráci se ZČU sestrojí funkční pikosatelit a navrhnu experimenty, které se na oběžné dráze zrealizují. Projekt také podpořilo město Plzeň.

Tato práce se zabývá prováděním modální analýzy kostry za pomoci MKP softwaru Abaqus. Předpokládá se provedení experimentu pro určení modálních charakteristik. Data budou sloužit k ověření numerického modelu. Nyní je materiál kostry hliník, do budoucna je plán návrhu kostry z kompozitního materiálu.

### 2 Modální analýza

Modální analýza je metoda pro identifikaci modálních vlastností zkoumaného objektu. U lineárních systémů nejsou modální veličiny ovlivněny vstupní veličinou (buzením) a v případě slabě tlumených systémů ani tlumením. Modální analýza na netlumeném a nebuzeném systému je popsána rovnicí:

$$\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{K}\mathbf{q}(t) = \mathbf{0}, \quad (1)$$

kde  $\mathbf{M}$  je matice hmotnosti,  $\mathbf{K}$  je matice tuhosti. Po dosazení předpokládaného řešení:

$$-\mathbf{q}(t) = \mathbf{v} \sin(\Omega t), \quad (2)$$

vykrácení časovou funkcí dostaneme:

$$(\mathbf{K} - \Omega^2 \mathbf{M})\mathbf{v} = \mathbf{0}. \quad (3)$$

Modální a spektrální matice jsou nadefinovány:

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \in \mathbf{R}^{n,n}, \quad (4)$$

$$\mathbf{\Lambda} = \text{diag}\{\Omega_i^2\}, \quad (5)$$

kde  $\mathbf{V}$  je modální matice,  $\mathbf{\Lambda}$  je spektrální matice a  $\Omega_i$  je vlastní úhlová frekvence [rad/s]=[s<sup>-1</sup>] pro  $i$ -tý vlastní tvar kmitu. Výsledek modální analýzy diskrétních matematických modelů se slabým tlumením a se symetrickými maticemi  $\mathbf{M}$ ,  $\mathbf{K}$  jsou modální a spektrální matice splňující podmínky:

$$\mathbf{V}^T \mathbf{M} \mathbf{V} = \mathbf{I}, \quad (6)$$

$$\mathbf{V}^T \mathbf{K} \mathbf{V} = \mathbf{\Lambda}, \quad (7)$$

kde  $\mathbf{I}$  je jednotková matice.

---

<sup>1</sup> student bakalářského studijního programu Počítačové modelování v technice, obor Výpočty a design, e-mail: jlinhart@students.zcu.cz

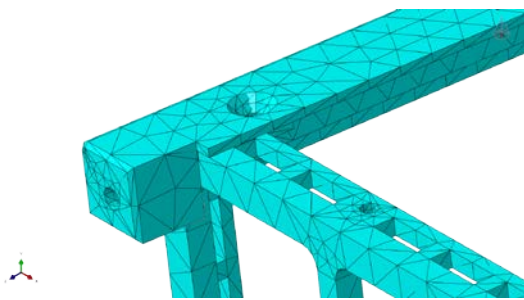
### 3 Výsledky

Pomocí MKP programu Abaqus je provedena modální analýza kostry o tvaru krychle o hraně 10 cm. Výpočet proběhl pro frekvence mezi 1-2000 Hz. Materiál kostry je hliník s konstantami: hustota  $\rho = 2700 \text{ kg}\cdot\text{m}^{-3}$ , modul pružnosti  $E = 71 \text{ GPa}$  a Poissonova konstanta  $\nu = 0,34$ . CAD model byl převzatý. Jelikož je model zadaný v centimetrech, materiálové konstanty jsou převedeny na:  $\rho = 2700 \times 10^{-6} \text{ kg}\cdot\text{cm}^{-3}$ ,  $E = 71 \times 10^7 \text{ kg}\cdot\text{cm}^{-1}\cdot\text{s}^{-2}$ . Typ prvků při síťování, které lze vidět na Obrázku 1, je tetrahedron. Počet těchto prvků na modelu je 125 428. Jejich přibližná globální velikost je 0,4 cm. Kostra je složena z 9 částí a 4 šroubů s matkami, které lze vidět na Obrázku 2. Spodní část je určena pro baterie, v horní části se budou nacházet plošné spoje. V reálu jsou tyto části spojeny šrouby. Pro výpočet jsou jednotlivé části spojeny pomocí vazeb typu Tie Constraints. Výsledné vlastní frekvence jsou zahrnuty v Tabulce 1 a tvary kmitu lze vidět na Obrázku 3 resp. Obrázku 4. Frekvence  $f$  je odvozena ze vztahu:

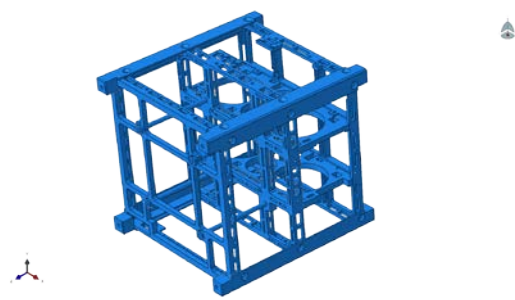
$$\Omega = 2 * \pi * f, \quad (8)$$

$i$	1	2	3	4	5	6	7
$f$ [Hz]	928,93	986,68	1085,8	1233,5	1320,5	1719,2	1905

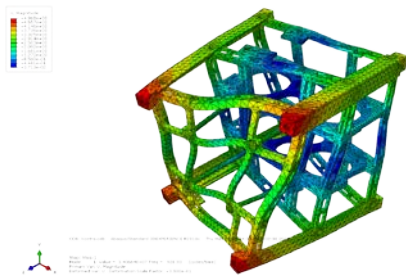
**Tabulka 1:** Výsledné vlastní frekvence.



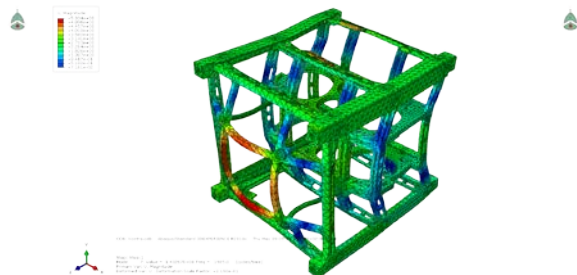
**Obrázek 1:** Detail sítě modelu.



**Obrázek 2:** Sjednocený model.



**Obrázek 3:** Vlastní tvar pro  $i = 1$ .



**Obrázek 4:** Vlastní tvar pro  $i = 7$ .

### 4 Závěr

Podářilo se navrhnout a zrealizovat převod z CAD modelu do numerického modelu a určit modální charakteristiky. Věřohodnost numerického modelu bude dále ověřena na problému jednostranně vetknutého nosníku, kde je známo analytické řešení vlastních frekvencí pomocí Euler-Bernoulliho teorie.

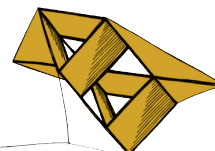
### Literatura

Dupal, J. (2012) *Mechanika 3*, Plzeň.



*Sekce*

**Informatika, kybernetika**



# Metody strojové klasifikace pro výběr optimálního estimátoru křivosti

Filip Hácha<sup>1</sup>

## 1 Úvod

Tato práce se zabývá zkoumáním několika různých přístupů z oblasti strojového učení pro určení optimálního estimátoru křivosti trojúhelníkových sítí na základě jejich vlastností. V práci jsou prezentovány výsledky dosažené za pomoci různých klasifikačních algoritmů.

Při konstrukci klasifikátorů některých klasifikátorů byly dále využity optimalizační pro nalezení jeho optimální konfigurace, algoritmy pro normalizaci a analýza hlavních komponent pro redukci dimenze vstupních dat.

## 2 Křivost

Křivost je zajímavou vlastností ploch, neboť umožňuje rozpoznání rysů jako jsou údolí, roviny, konvexní, konkávní nebo sedlové tvary, jak ukazují Gatzke a Grimm (2006). Jak poukázali Váša et al. (2017), zatímco výpočet křivosti v bodě na hladkých površích běžně nepředstavuje obtížnou úlohu, odhad křivosti v bodech polygonálních sítí je obtížnější, neboť rozdílná tělesa mohou být reprezentovány stejnou trojúhelníkovou sítí.

Způsobů, jak odhadovat křivost původního tělesa na základě trojúhelníkové sítě, je více a pro různá data nám poskytují různé estimátory odhady s rozdílnou přesností. Výběr optimálního estimátoru pro konkrétní síť tedy můžeme označit za klasifikační úlohu.

## 3 Předchozí práce

Tato práce vychází z výzkumu, který v této oblasti provedli Váša et al. (2016), který se zabýval implementací jednotlivých estimátorů křivosti a ve kterém byl zároveň popsán klasifikátor založený na rozhodovacích stromech s jedním a se dvěma rozhodovacími uzly.

## 4 Metody klasifikace

Vstupními daty úlohy je množina trojúhelníkových sítí, která je parametrizována vektorem různých vlastností a chybou jednotlivých estimátorů na dané síti. Cílem úlohy je přiřadit sítím na základě vektoru sledovaných vlastností optimální estimátor křivosti.

Rozsahy hodnot sledovaných vlastností se od sebe řádově liší, proto je pro některé klasifikační algoritmy nutné provést jejich normalizaci. Prvním implementovaným přístupem je normalizace vlastností, jelikož je tato metoda ale velmi náchylná na odlehlá měření, byla později nahrazena normalizací směrodatnou odchylkou.

Pro některé některé způsoby klasifikace je počet sledovaných vlastností příliš vysoký a jelikož lze předpokládat, že některé z vlastností mohou být navzájem závislé, je možné provést

---

<sup>1</sup> student bakalářského studijního programu Aplikované vědy a informatika, obor Informatika, e-mail: ha-chaf@kiv.zcu.cz

jejich redukci. Pro tento účel byla implementována analýza hlavních komponent, která umožnila např. prohledání větší části prostoru možných řešení při konstrukci rozhodovacího stromu.

Práce porovnává úspěšnost klasifikace trojúhelníkových sítí s využitím klasifikátorů, které byly zkonstruovány za pomoci rozhodovacích stromů, neuronových sítí, algoritmu pro shlukování a lineární  $L^2$  aproximace. Pro konstrukci rozhodovacích stromů s více vrstvami byly vyzkoušeny optimalizační metody simulovaného žíhání a genetické algoritmy, které spolu s redukcí počtu sledovaných vlastností umožnily konstrukci stromů s větším počtem rozhodovacích uzlů

Úspěšnost jednotlivých klasifikátorů byla porovnána stejně jako ve výchozí práci pomocí relativní chyby vypočtené podle vzorce 1.

$$e_{rel} = \frac{e_i - e_{min}}{e_{min}} \quad (1)$$

## 5 Výsledky

Nejlepších výsledků bylo dosaženo zredukováním dimenze vstupních dat analýzou hlavních komponent a následnou konstrukcí binárního rozhodovacího stromu se třemi vrstvami rozhodovacích uzlů za využití optimalizace pomocí genetického algoritmu. Takto sestavený klasifikátor dosáhl relativní chyby  $e_{rel} = 0.59$ , což je zlepšení oproti klasifikátoru, který uvedli Váša et al. (2016), s relativní chybou  $e_{rel} = 0.86$ .

## 6 Závěr

V průběhu práce se podařilo dosáhnout výrazného zlepšení úspěšnosti klasifikace trojúhelníkových sítí oproti výchozímu výzkumu, můžeme však očekávat, že u některých použitých klasifikačních algoritmů nebyl zcela naplněn jejich potenciál a je zde tedy prostor k dalšímu výzkumu.

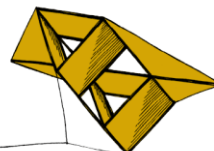
V budoucí práci je například možné hlouběji prozkoumat možnost použití neuronových sítí a také sofistikovanější metody pro konstrukci rozhodovacích stromů, jako například využití TDIDT (Top down induction of decision trees) algoritmů.

## Poděkování

Děkuji panu doc. Ing. Liborovi Vášovi, Ph.D. za ochotu při vedení bakalářské práce a rady s jejím vypracováním. Dále děkuji panu doc. Ing. Pavlovi Královi, Ph.D. za poskytnutí konzultací k problematice neuronových sítí.

## Literatura

- D. Gatzke, T. a Grimm, C. (2006) *Estimating Curvature on Triangular Meshes* International Journal of Shape Modeling, Volume 12, pp. 1-28.
- Váša, L. a Kühnert, T. a Brunnett G. (2017) *Multivariate analysis of curvature estimators* Computer-Aided Design and Applications, Volume 14, pp. 58-69.
- Váša, L. a Vaneček, P. a Prantl, M. a Skorkovská, V. a Martínek, P. a Kolingerová, I. (2016) *Mesh Statistics for Robust Curvature Estimation* Eurographics Symposium on Geometry Processing, Volumen 35, pp. 271-280.



## Detekce tématu dokumentu

Jan Kandyba<sup>1</sup>

### 1 Úvod

Příspěvek se zabývá problematikou zpracování přirozeného jazyka, konkrétně klasifikací dokumentů. Klasifikace dokumentů na základě jejich obsahu je v dnešní době velice používaná, aniž bychom si to příliš uvědomovali – například ideálním příkladem je každodenní klasifikace spamových emailů. V tomto příspěvku byla testována efektivnost vybraných klasifikátorů na datech poskytnutých od Ústavu pro jazyk český Akademie věd ČR. Mezi vybrané algoritmy patří Support Vector Classification, lineární diskriminační analýza a aplikace neuronových sítí.

Data od ÚJČ tvoří ručně anotované rozhovory mezi dotazujícími a jazykovými poradci – dotazující zavolá na jazykovou poradnu s nějakým dotazem a poradci se jej snaží zodpovědět. Data byla anotována ručně, aby se vyhnulo problémům z automatického přepisu. Tento dotaz může být různého typu, existuje však několik častých témat, na která se dotazující ptají nejvíce. Tato práce navazuje na projekt NAKI, jehož cílem je vytvořit semi-automatizovaný systém, který usnadní vyhledávání a kategorizaci těchto dotazů. Strom tříd, do kterých poradci ÚJČ tyto dotazy klasifikovali je však poměrně rozsáhlý, a ne pro každou třídu zde bylo dost dat, aby se klasifikátor natrénoval správně. Bylo tedy nutno udělat shlukování menších tříd tak, aby je bylo možné dobře rozpoznat.

### 2 Experimentální část

Abychom mohli provést automatickou klasifikaci, je nutno data parametrizovat do číselné podoby. Způsobů, jakými lze text parametrizovat, existuje několik. V této práci byla provedena parametrizace pomocí 2 přístupů: TFIDF a doc2vec. TFIDF se skládá ze 2 částí – TF a IDF. TF označuje četnost slova v dokumentu (tato hodnota se často normalizuje na celkový počet slov v dokumentu) a IDF složka reprezentuje „důležitost“ slova – čím častěji se slovo vyskytuje napříč všemi dokumenty, tím méně je důležité, resp. dává menší informaci o tématu daného dokumentu. Doc2vec parametrizace vychází z word2vec parametrizace, která zohledňuje kontext daných slov. Doc2vec navíc přidává číslo dokumentu k dané parametrizaci slov. Bylo zjištěno, že doc2vec parametrizace není na tuto úlohu vhodná zejména z důvodu nízkého množství trénovacích dat. Pro finální výsledky byla tedy použita parametrizace TFIDF.

Po parametrizaci na číselné hodnoty byly provedeny experimenty s různými nastaveními parametrů metod a byly porovnány jejich výsledky. Po prvních experimentech se pohybovala úspěšnost klasifikace všech metod okolo 90 %. Vysoká úspěšnost klasifikace byla zapříčiněna přítomností značně nevyvážené třídy „balast“, která obsahovala řádově více dat, než ostatní třídy – při trénování se tak klasifikátor natrénoval především pro tuto třídu, ale nebyl schopen klasifikovat správně do ostatních tříd. Navíc většina dat pocházela právě z třídy balast, proto byla úspěšnost takto vysoká. Tento problém byl vyřešen tak, že při každém trénování byla náhodně načtena pouze část dat ze třídy balast – tak, aby data byla vzájemně vyvážená.

---

<sup>1</sup> student bakalářského studijního programu Inženýrská informatika, obor Systémy pro identifikaci, bezpečnost a komunikaci, e-mail: kandybaj@students.zcu.cz

Po vyvážení tříd klesla úspěšnost klasifikace na necelých 70 %. Dále byly prováděny experimenty za účelem zlepšení přesnosti klasifikátorů. Před parametrizací dat se přidala lemmatizace, která z původních slov vytvořila základní tvar – například slovo „lepší“ se lemmatizuje na „dobrý“. Lemmatizace mírně zlepšila výsledky, avšak nikterak zásadně. Bylo zjištěno, že největší vliv na správnou klasifikaci má správná volba příznaků – tedy volba parametrů TFIDF. Pokud jsou zvoleny nevhodně, podepíše se to i na úspěšnosti klasifikace.

Vybrané klasifikátory tvoří klasifikace pomocí podpůrných vektorů SVC, neuronové sítě (ANN) a lineární diskriminační analýza (LDA), která je též použita pro redukci dimenze příznaků a následně se klasifikovalo pomocí SVC.

SVC klasifikátor využívá podpůrných vektorů, tedy bodů ležících na okraji každé třídy. Jelikož třídy nejsou lineárně separabilní, využívá SVC jádrový trik, kterým transformuje body do vyššího prostoru, kde již lineárně separabilní jsou.

Princip LDA spočívá v nalezení lineární kombinace příznaků, které oddělují dané třídy, nebo které mají mezi sebou společné. LDA lze využít jako samostatný klasifikátor, nebo jej lze využít pro snížení dimenze příznaků před následnou klasifikací.

Předchozí metody využívají efektivních přístupů, avšak nenabízí velké možnosti volitelných parametrů. Neuronové sítě nabízí značnou svobodu v jejich návrhu – počet vrstev, počet neuronů v jednotlivých vrstvách, volba aktivační funkce atd. Výsledná síť obsahovala 3 vrstvy: vstupní, 1 skrytou s aktivační funkcí *sigmoid* a výstupní s funkcí *softmax*.

Přestože se s aktivační funkcí *sigmoid* pojí problém klesající gradient, osvědčila se jako nejefektivnější aktivační funkce. Propojení mezi vstupní a skrytou vrstvou nebylo úplné, jelikož se mezi ně vložila „vrstva“ v Pythonu označována jako *dropout*, která náhodně vyřadila určité množství neuronů ze vstupní vrstvy – skrytá vrstva tak při trénování vidí pouhou část dat. Neuronová síť se tak nemůže spoléhat na všechna data, což vede ke zlepšení klasifikace.

### 3 Závěr

Nejlepší výsledky klasifikace lze dosáhnout při správné volbě příznaků. Pokud jsou zvoleny nevhodně, bude špatná i úspěšnost klasifikace. Volba správných příznaků není jednoznačná, výsledky v následující tabulce tak nemusí být nejlepší možné, kterých lze při těchto datech dosáhnout.

Zlepšení výsledků klasifikace může zlepšit předzpracování dat, které nepřímo souvisí se zvolenou parametrizační metodou. Nejvyšší přesnosti dosáhl klasifikátor SVC.

Typ Metody	Průměrné skóre	Minimální skóre	Maximální skóre
SVC	66,88 %	63,37 %	71,04 %
LDA	63,22 %	59,16 %	66,09 %
LDA-SVC	62,85 %	58,66 %	65,59 %
ANN	64,18 %	59,65 %	66,83 %

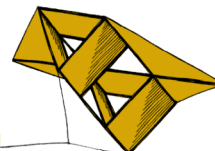
**Tabulka 1:** Výsledky klasifikace pro různé metody; použitá metrika: přesnost

### Literatura

Hořejš Jiří, Kubát Miroslav, Lažanský Jiří, Mařík Vladimír, Štěpánek Petr, Štěpánková Olga, Zdráhal Zdeněk. "Umělá inteligence", Akademie věd České republiky, 1993.

Adrian A. Hopgood. "Intelligent Systems for Engineers and Scientists", CRC PR INC., 2011.

Zbyněk Zajíc, Lucie Zajícová, Josef V. Psutka, Petr Salajka, Jaromír Novotný, Aleš Pražák, Luděk Müller. "First Insight into the Processing of the Language Consulting Center Data", SPECOM 2018, p. 778-787, Springer, 2018.



## Elektronický zápisník pro diabetiky

Aneta Koldovská<sup>1</sup>

### 1 Úvod

Cukrovka je onemocnění slinivky břišní, při kterém je nutné dodržovat stálý režim stravování a pravidelný přísun potravin. Jak popsal Edelsberg (2009), je nutné si v rámci kompenzace této nemoci vést přehled o hladině cukru v krvi, která je u diabetika proměnlivější než u zdravého člověka. V dnešní době jsou způsoby vedení těchto denních záznamů mnohdy velmi nepřehledné, jelikož jsou hlavně v papírové podobě (viz obrázek 1). Právě proto jsem se rozhodla vytvořit elektronický zápisník, který by diabetikům zápis těchto údajů usnadnil (viz obrázek 2).

V rámci vývoje jsem již na střední škole začala spolupracovat s Fakultní nemocnicí v Plzni a absolvovala stáž na Diabetologické ambulanci u paní doktorky MUDr. Silvie Lacigové, Phd. Mohla jsem tak lépe určit parametry mého programu, aby co nejvíce vyhovoval potřebám pacientů.

### 2 Popis aplikace

Program například umí automaticky vypočítat sacharidové jednotky, jelikož jak uvádí Diabetická asociace ČR (2014), sacharidy jsou stěžejní při kompenzaci tohoto onemocnění, a tak si diabetici musí v rámci příjmu potravy množství přijatých sacharidů hlídat. Výpočet využívá především jednu ze dvou konstant (10 nebo 12), kterou si diabetik může zvolit v sekci „Informace o uživateli“. Dále výpočet využívá hodnoty sacharidů v potravinách. Základní tabulka těchto sacharidů je také obsažena v mém programu. Diabetik tedy již nemusí konkrétní hodnoty zdlouhavě hledat.

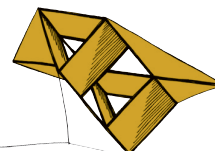
Aplikace zároveň umí diabetika upozornit na nebezpečné zvýšení či snížení hladiny cukru v krvi, což v některých případech může být pro pacienty s tímto onemocněním až život ohrožující. Dále si uživatel může ke svému záznamu přidat i bolusy (množství inzulínu, který si diabetik dodává ke každému jídlu) a množství ketonů (škodlivých látek) v moči, nebo si formou poznámky záznam doplnit například o popis konkrétního jídla, které jedl. V rámci studia na ZČU jsem program dále rozvíjela jako semestrální práci v předmětu KIV/UUR (Úvod do uživatelských rozhraní). Zápisník jsem optimalizovala hlavně z uživatelského hlediska (přidala jsem nové grafické prvky, například kalendář, ze kterého si uživatel zvolí datum při vyplňování nového záznamu). Hlavní změnou je možnost editace tabulky Zápisníku.

Program je určen pro stolní počítače a notebooky a byl testován na operačních systémech Windows 10 a Linux Debian 9. Aplikaci jsem vyvíjela v programovacím jazyce JavaFX, za použití vývojového prostředí IntelliJ Idea. Zápisník je v dnešní době aktivně využíván několika diabetiky.

---

<sup>1</sup>studentka bakalářského studijního programu Inženýrská informatika, obor Informatika,  
e-mail: anetkold@students.zcu.cz





## Full-textové vyhledávání s podporou porozumění textu dotazu

Adam Mištera<sup>1</sup>

### 1 Úvod

Cílem práce je prostudovat dostupné metody sémantické reprezentace textu a na základě této studie navrhnout a implementovat full-textové vyhledávání s podporou porozumění textu dotazu, které se následně integruje do předem zvoleného full-textového vyhledávače *Apache Solr*. Výstupem práce bude zhodnocení získaných výsledků, respektive analýza účinnosti daných metod vzhledem ke zvýšení kvality vyhledávání ve vybrané datové kolekci.

### 2 Sémantická reprezentace

Pro vytvoření sémantické reprezentace textu byly zvoleny dva modely *Word2vec* a *FastText*. **Word2vec** (Mikolov et al. (2013)) je skupina modelů založených na umělých neuronových sítích určených pro vytvoření slovních vektorů (anglicky *word embedding*). Vstupem modelu *Word2vec* je obsáhlý soubor textů vybraného jazyka, jinak také zvaný jazykový korpus. Příkladem vhodného datového korpusu je internetová encyklopedie *Wikipedia*, která v anglické mutaci obsahuje několik milionů různých článků. Proces vytvoření slovních vektorů spočívá v převedení frází, popřípadě slov jazykového korpusu na vektory umístěné ve vektorovém prostoru většinou čítajícím stovky dimenzí. Jednotlivým slovům poté odpovídá vždy jeden vektor z tohoto prostoru. Slovní vektory sdílející podobný či blízký význam jsou zpravidla umístěny blízko sebe. Naproti tomu nesouvisející slova či slova opačného významu jsou ve vektorovém prostoru daleko od sebe.

**FastText** (Ryan (2016)) je rozšíření modelu *Word2vec* navržené v roce 2016 firmou *Facebook* provozující největší sociální síť na světě. Na rozdíl od modelu *Word2vec*, který považuje každé slovo v korpusu za atomickou jednotku, *FastText* považuje slovo za množinu několika n-gramů. Pojem n-gram je definován jako řetězec prvků (jako například slova či písmena), které se objevují v delší posloupnosti. Například trigramy slova *jablko* jsou *jab*, *abl*, *blk* a *lko*. Výsledný slovní vektor slova *jablko* je poté součtem vektorů jednotlivých n-gramů.

### 3 Apache Solr

*Apache Solr* (Grainger a Potter (2014)) je open-source platforma napsaná v jazyce *Java*, která byla navržena a optimalizována pro full-textové prohledávání a indexování rozměrných textových dat.

Na rozdíl od většiny běžných databázových systémů, kde je k jednotlivým dokumentům přiřazeno pole slov, která jsou součástí dokumentu, *Apache Solr* používá odlišnou metodu, takzvaný **invertovaný index** (Grainger a Potter (2014)). Invertovaný index ke každému slovu nebo výrazu obsaženému v korpusu mapuje všechny dokumenty, ve kterých je obsažen.

---

<sup>1</sup> student bakalářského studijního programu Inženýrská informatika, obor Informatika, e-mail: amistera@students.zcu.cz



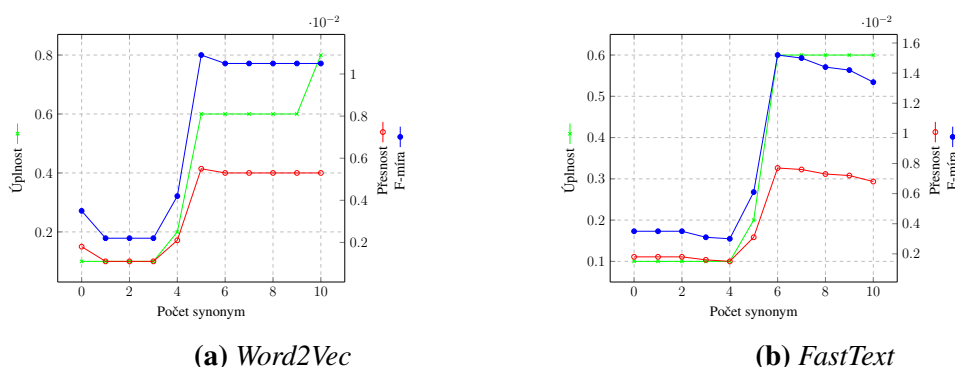
## 4 Řešení a experimenty

Řešením bylo sestavení rozšířeného dotazu, který vznikne přidáním synonym získaných z natrénovaného modelu k textu dotazu. Uvažujme například, že se uživatel pokouší vyhledat následující dotaz: „Nehody v zaměstnání“. Rozšířený dotaz sestavíme přidáním slov *havárie* a *neštěstí* ke slovu *nehody*. Do druhé části dotazu, konkrétně ke slovu *zaměstnání*, přidáme výrazy *povolání* a *podnikání*.

Funkčnost navrženého řešení byla ověřena na datové kolekci **CLEF**<sup>1</sup> obsahující více než 80 000 novinových článků z periodik *Mladá Fronta* a *Lidové noviny*. K těm bylo přiřazeno 50 různých témat, které byla použita pro ověření funkčnosti navrženého řešení. Při vyhodnocení výsledků kvality námi navrženého systému jsme sledovali tři hlavní míry, které jsou v praxi běžně používané především v oblasti získávání informací. Jedná se o *přesnost*, *úplnost* a *F-míru*.

### 4.1 Téma *Ohrožené druhy*

Na obrázku 1 můžeme vidět výsledky pro téma *425-AH Ohrožené druhy* současně pro oba modely *Word2vec* i *FastText*. Na první pohled je zcela zřejmé, že došlo k velmi výraznému zlepšení *přesnosti*, *úplnosti* i *F-míry*.



Obrázek 1: Výsledky pro téma *425-AH*

## 5 Závěr

S použitím modelů *Word2vec* a *FastText* se ve všech experimentech podařilo zvýšit hodnoty *úplnosti* vyhledávání pouze s mírným snížením *přesnosti*. K dalšímu navýšení hodnot *přesnosti* by téměř jistě došlo za předpokladu, že bychom použili pro dané výrazy výhradně synonyma.

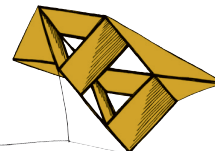
## Literatura

Grainger, T. a Potter, T. (2014) *Solr in Action*. New York, Shelter Island.

Mikolov, T., Chen, K., Corrado, G. a Dean, J. (2013) *Efficient Estimation of Word Representations in Vector Space*. Available from: *arXiv preprint arXiv:1301.3781*.

Ryan, K. J. (2016) *Facebook's New Open Source Software Can Learn 1 Billion Words in 10 Minutes*.

<sup>1</sup><http://catalog.elda.org/en-us/repository/browse/ELRA-E0036/>



# Systém pro rozesílání cvičných phishingových zpráv

Martin Šebela<sup>1</sup>

## 1 Úvod

S tím, jak se internet stal standardem běžného života dnešní společnosti, došlo k tomu, že stejné médium začali využívat i lidé, jejichž cílem je z méně pozorných uživatelů získat důvěrné informace (hesla, čísla platebních karet apod.), a ty následně zneužít ve svůj prospěch. Typickým příkladem je phishing (např. podvodné e-maily) a s ním související metody sociálního inženýrství, kterých útočník využívá.

Cílem této bakalářské práce tak bylo vytvořit systém, který umožní rozesílat cvičné phishingové zprávy (včetně odkazů na cvičné podvodné webové stránky), na základě kterých budou dobrovolně registrovaní uživatelé schopni lépe odhalovat reálný phishing. Uživatelé mohou z těchto cvičných podvodných e-mailů zjistit, na co se mají v elektronické komunikaci zaměřit, jakých metod mohou útočníci využívat, a především se v případě neúspěchu i poučit – systém uživateli poskytne zpětnou vazbu, čeho si měl na podvodném e-mailu všimnout a podle jakých indicií mohl prohlásit, že se jedná o phishing.

## 2 Vytvořený systém Phishingator

Systém byl vytvářen tak, aby nesloužil pouze administrátorům k rozesílání cvičného phishingu, ale i běžným uživatelům, čímž se liší od všech ostatních, existujících řešení. Existující řešení navíc neobsahují funkce požadované univerzitou, přičemž některá z nich jsou i finančně nákladná. V minulosti byl na univerzitě cvičný phishing rozesílán ve spolupráci s externí organizací *CESNET*.

Administrátor je schopen v naprogramovaném systému během několika minut vytvořit novou phishingovou kampaň, která se skládá z rozesílaného podvodného e-mailu, podvodné webové stránky, jejíž odkaz bude v e-mailu uveden, dále pak z akce, která se stane po odeslání formuláře na podvodné stránce a konečně také ze seznamu příjemců (a dalších parametrů jako data spuštění a ukončení kampaně apod.). Všechny zmíněné fragmenty phishingové kampaně lze v systému také vytvářet a dále spravovat (tj. podvodné e-maily, indicie k rozpoznání phishingu u daného e-mailu a podvodné webové stránky).

Po spuštění phishingové kampaně dochází ke sledování příjemců, a sice tak, že jsou na podvodné stránce zaznamenávány všechny akce uživatelů (tj. zdali stránku navštívili, zdali vyplnili platné či neplatné přihlašovací údaje, nebo zdali na podvodný e-mail vůbec nereagovali). Všechna zaznamenaná data jsou systémem zpracována a přehledně zobrazena do několika grafů a tabulek, přičemž získaná data je také možné exportovat do formátu CSV (*Comma-separated values*).

Běžní uživatelé (kdokoliv s univerzitním kontem) se do systému mohou přihlásit a zároveň se dobrovolně registrovat k odebírání cvičných phishingových zpráv (včetně možnosti nastavit

---

<sup>1</sup> student bakalářského studijního programu Inženýrská informatika, obor Informatika,  
e-mail: msebela@students.zcu.cz

si případný limit). Uživatelé si mohou všechny přijaté phishingové e-maily zpětně prohlédnout, a to včetně indicií (označených pasáží v textu včetně popisu), na základě kterých bylo možné phishing rozpoznat (viz obr. 1).

moje reakce:  
**zadání platných údajů**  
odesláno 16. 4. 2019 1:15

---

Od: **Stravovací viceprezident SKM** <vice@zcu.cz>  
Předmět: Obědové stipendium pro zaměstnanci  
Komu: msebela@students.zcu.cz

Vážený strávniku msebela,  
díky grantu EU 167/7669077-6429CZ je možné se přihlásit k bezplatným **objedům**.

Přihlásit se zde: **https://zcu.webkdc.cz**

Časový limit pro **přihlášení** je 22. 5. 2019

Bc. et. Bc. Jan Hledový  
**Stravovací viceprezident**  
vice@skam.zcu.cz  
+555 123 646 888  
SKM, ZČU ve v Plzni

---

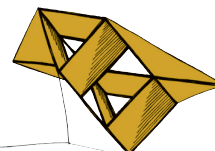
<b>1. indicie</b> Podivná funkce Neexistující a z názvu i podezřelá funkce <a href="#">Zruš označení</a>	<b>3. indicie</b> Jazyk zprávy Viceprezident firmy zajišťující stravování by měl vědět jak napsat hlavní jídlo dne. <a href="#">Zruš označení</a>	<b>4. indicie</b> Jazyk zprávy Další překlep <a href="#">Zruš označení</a>
<b>2. indicie</b> Podezřelá doména Doména zcu.webkdc.cz nepatří ZČU <a href="#">Zruš označení</a>		<b>5. indicie</b> Neexistující funkce Pro poměry ZČU poměrně podivná funkce <a href="#">Zruš označení</a>

**Obrázek 1:** Jeden ze cvičných phishingových e-mailů včetně seznamu indicií (zakroužkovaných pasáží v textu včetně souvisejícího popisu v dolní části obrázku) tak, jak si jej může uživatel zobrazit ve vytvořeném systému (tento e-mail byl zároveň zasílán vybraným uživatelům v rámci testování bakalářské práce, přičemž v horní části obrázku je uživateli zobrazeno, jak na e-mail, popř. na podvodnou webovou stránku reagoval)

System byl naprogramován ve skriptovacím jazyce PHP s využitím architektury MVC (*Model-View-Controller*), data se ukládají do MySQL databáze a celý systém je propojen s univerzitním LDAP (*Lightweight Directory Access Protocol*) a autentizační službou *WebAuth*.

### 3 Závěr

Vytvořený systém byl v rámci bakalářské práce pilotně otestován zaměstnanci i studenty a nasazen na server *Západočeské univerzity v Plzni*. Systém je tak dostupný všem uživatelům univerzity, a to na oficiální URL adrese <https://phishingator.zcu.cz>. Systém bude dále využíván ke vzdělávání uživatelů univerzity v oblasti phishingu a dále vylepšován o nové funkce. Systém byl zároveň navržen jako modulární a mohl by tak být nasazen i na jiné univerzitě.



## Autonomní navigace kolového robota

Josef Švec<sup>1</sup>

### 1 Úvod

Cílem této práce bylo sestavit funkční řídicí systém tak, aby kolový robot dojel na místo určené v mapě, kterou si sám vytváří během jízdy, a sledoval vytyčenou trajektorii co nejlépe. Ke splnění tohoto cíle bylo nutné vyřešit několik dílčích úkolů. Nejprve bylo potřeba zkompletovat robotickou platformu a vyřešit napájení všech součástí. Dále bylo nutné řešit především řízení motorů tak, aby se otáčely požadovanou úhlovou rychlostí. V dalším kroku bylo potřeba vytvořit model řízeného robota a následně s využitím frameworku ROS (Mahtani et al. (2018)) sestavit celý systém tak, aby plnil zadaný úkol.

### 2 Řízení motorů

Bylo nutné zajistit, aby se motory otáčely požadovanou úhlovou rychlostí. Pro řízení stejnosměrných (DC) motorů se nejčastěji využívají PID regulátory. Jedná se o zpětnovazební regulátory, které využívají měření aktuální úhlové rychlosti pomocí enkodérů a podle toho upravují velikost akčního zásahu. Tento přístup zásadně ovlivňuje přesnost a kvalitu regulace (Astrom a Murray (2008)). Řízení DC motorů kolového robota v této práci bylo prováděno mikrořadičem, takže byla využita diskretní verze PI regulátoru, jako speciálního případu spojitého PID regulátoru. Derivační složka nebyla využita, protože by docházelo k zesílení šumu měření.

### 3 Model diferenciálně řízeného robota

Diferenciálně řízený robot je takový, kde se nezávisle pohybuje levá a pravá strana podvozku. Zatáčení se provádí smykem a robot je schopný se otočit na místě. Bylo nutné popsat přesný kinematický model, který stanoví pozici robota v prostoru (rovině) na základě časových průběhů rychlostí kol na jeho levé a pravé straně. Řízení kolových robotů je většinou řešeno kaskádně. Tento přístup zvyšuje celkovou přesnost regulace. Nižší stupeň řídí požadovanou úhlovou rychlost otáčení kol a vyšší stupeň řízení zajišťuje regulaci na požadovanou polohu a orientaci.

### 4 ROS

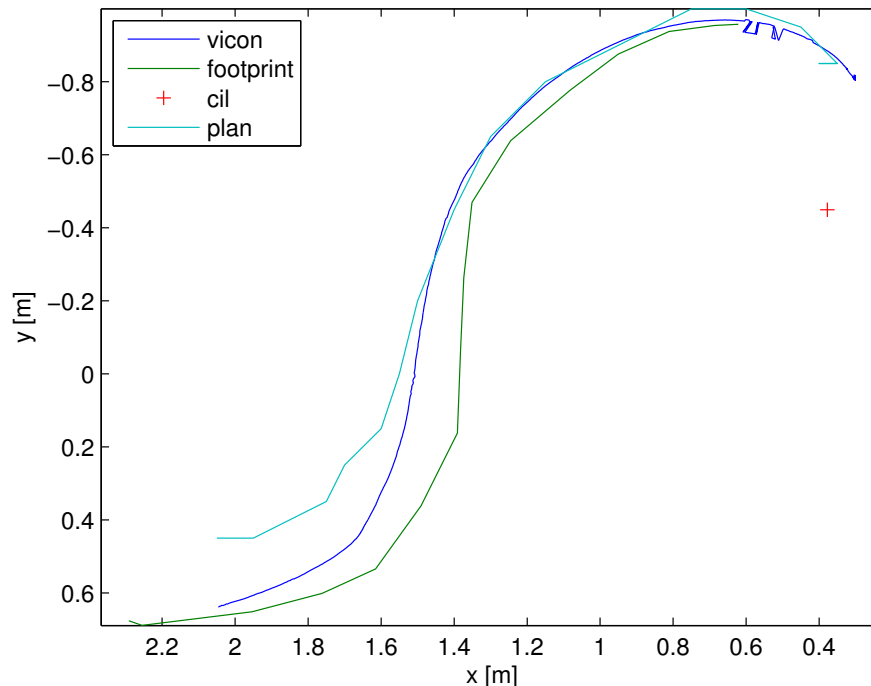
Softwarový rámec ROS (Robot Operating System) je v dnešní době velmi rozšířeným nástrojem pro vývoj robotických systémů. Jeho hlavní výhodou je přenositelnost mezi různými HW platformami. ROS byl využit z toho důvodu, že obsahuje veliké množství hotových balíčků vhodných pro řešení této úlohy (Navigation stack, Gmapping (Grisetti et al. (2007))). Pro zprovoznění tohoto výkonného nástroje je potřeba znát přesný kinematický model robota. Dále je nutné používat lidar, aby se robot mohl vyhnout případným nepředvídatelným překážkám.

---

<sup>1</sup> student bakalářského studijního programu, obor Kybernetika, e-mail: svecpepik@gmail.com

## 5 Experimentální testování

Měření slouží k porovnání naplánované a realizované trajektorie a dále slouží k ověření dosažení zadaného cíle. Robot byl postaven před překážku (papírová krabice) a cíl byl umístěn až za ní. Pro změření skutečně realizované trajektorie byl použit systém Vicon. Naplánovaná trajektorie lze zobrazit ve vizualizačním nástroji Rviz (součást ROS).



**Obrázek 1:** Uskutečněná trajektorie okolo překážky

Robot našel trajektorii s nejmenší cenou a následně trajektorii realizoval. Cílový bod byl dosažen se zadanou přesností. Překážka nebyla porušena.

## 6 Závěr

Výsledkem této práce je kolový robot, který dojde na místo určení v mapě, kterou si sám vytváří pomocí algoritmu Gmapping z lidarových a odometrických dat. Pokud se v naplánované trajektorii objeví překážka, najde si robot alternativní trasu nebo se zastaví.

## Literatura

- Astrom, Karl J., Murray Richard M. (2008) (. *Feedback systems: an introduction for scientists and engineers*, Princeton: Princeton University Press, c2008. ISBN 978-0691135762.
- Grisetti, G., Stachniss, C., Burgard, W. (2007) Improved Techniques for Grid Mapping with Rao-Blackwellized Particle Filters. *IEEE Transactions on Robotics*, Volume: 23 , pp. 34–46.
- Mahtani, Sanchez, Fernandez, Martinez, Joseph (2018) *ROS Programming: Building Powerful Robots*, Packt Publishing Ltd. ISBN 978-1-78862-743-6.