



MASARYKOVA UNIVERZITA  
FAKULTA INFORMATIKY

## **DPRPO, Posterová soutěž**

19. 9. 2014

abstrakty přihlášených posterů



## **Influence of levodopa on brain network**

E. Bujnošková, J. Fousek, N. Elfmarková, M. Gajdoš, M. Mračková, I. Rektorová

Network analysis is a tool to evaluate brain connectivity structure. We show two ways how to define functional network — regions of interest specified by an expert, or created by cortex parcellation following anatomical AAL atlas — and analyze its topology. We used resting state fMRI data from 14 cognitively normal Parkinson's disease (PD) patients in their on and off medication states. We hypothesize that the medication reduces disconnectivity and causes changes in motor areas and different resting-state networks. We explored network densities in a broad range, binarized matrices and described the topology by degree, edge weight, path length, clustering coefficient, efficiency, and betweenness centrality. Differences in these network descriptors were statistically evaluated by paired Wilcoxon test. When examining network created by expert knowledge, we observed increase in global efficiency in ON state. Changes towards improved network topology were seen in all subnetworks (except insular) in ON state compared to OFF medication state. The approach using AAL atlas did not reveal any significant difference on global level, but the increase of connectivity caused by medication was observed in motor areas, further changes were registered in hippocampal regions and middle frontal gyrus. These findings partially support those from expert knowledge paradigm but both are influenced by small sample size. Study with increased cohort size and comparison with healthy controls will follow.

## Reflected Attacks Abusing Honeypots

Martin Husák, Martin Vizváry

We present the observation of distributed denial-of-service attacks that use reflection of the flooding traffic off reflectors. This type of attack was used in massive attacks against internet infrastructure of Czech Republic in March, 2013. The network of Masaryk University was abused as a reflector. The reflection of the traffic from the honeypots in our network caused false positive security incident detection and helped the attacker in reflecting more traffic. The attack is produced by botnets, networks of infected computers. Botnet is instructed to send TCP SYN packets to random hosts in the network at the same time. The source IP address is counterfeited so the responses floods the victim. Honeypots, which are by default set to accept any incoming network connection, unintentionally amplified the effect of reflection. We present an analysis of the attack from the point of view of honeypots and show the risks of having honeypots respond to any incoming traffic. While the network of Masaryk University in general reflected around 5 % of incoming traffic, the honeypots alone reflected 16 % of the traffic. A report from another university reported reflection 93 % of traffic incoming to honeypots. We also discuss the conclusions and lessons learned from the observed attack.

## Cloud-based Security Research Testbed: A DDoS Use Case

Tomáš Jirsík, Martin Husák, Zdeněk Eichler, Pavel Čeleda

Introducing new services and network management measures represents a digitálnícult task, since it is not easy to make this change without proper testing. Testing in a real environment, however, is not a suitable solution, as we need to maintain already running services and network configuration. A network simulation is a possible solution to this problem. Since there are various types of network topologies, services, and use-cases to test, a simulation testbed should be able to emulate the whole infrastructure, and capture the relevant network properties. Moreover, it should gain full control over all emulated activities. In our work, we focus on network security simulation and management, as cyber attacks have become ubiquitous.

Cybernetic Proving Ground (CPG) is a testbed designed especially for a network security management and simulation. It is a scalable, universal solution which is designed for deployment in clouds managed by OpenNebula. CPG provides an assisted generic network topologies emulation and an isolated virtual environments facilitation, which is used for controlled attack analysis. Moreover, it can serve as a cyber security training tool. The outputs can be used for research into new detection methods and for network management.

The purpose of this poster is to demonstrate an example of a network attack simulation in CPG. The demonstration focuses on a facilitation of network topology emulation, retaining main attack characteristics and real-world aspects of the simulation. We reproduced the Distributed Denial of Service (DDoS) attack against the internet infrastructure of the Czech Republic that took place in March, 2013. To reach our goal we had to specify a scenario based on the observed attacks, set up the testbed and run the experiment described in the scenario.

## **Math Indexer and Searcher Web Interface: Towards Fulfillment of Mathematicians' Information Needs**

Martin Líška, Petr Sojka, Michal Růžička

Skupina Math Information Retrieval (MIR) na Fakultě informatiky Masarykovy univerzity se zabývá výzkumem a vývojem nástrojů specifických pro matematický obsah (například formule) pro plnotextové vyhledávání v digitálních knihovnách odborných textů.

Předkládaný plakát shrnuje naši motivaci pro implementaci takového systému, seznamuje čtenáře se schématem principu práce našeho vyhledávacího systému Math Indexer and Searcher (MIaS), podpůrných nástrojů pro zpracování dat a zaměřuje se zejména na současnou podobu webového rozhraní systému MIaS (WebMIaS), která zpřístupňuje funkce systému koncovému uživateli.

WebMIaS poskytuje uživateli rozhraní pro dotazování pomocí klíčových slov kombinovaných s matematickými výrazy zapsanými v jazyce LaTeX hojně využívaným matematiky, nebo v XML jazyce MathML, který je často využíván jako univerzální formát pro výměnu dat mezi matematickým software a pro publikování na webových stránkách.

Pro pohodlí koncových uživatelů je zapisovaný dotaz v LaTeXu po každém novém znaku vizualizován, což poskytuje bezprostřední zpětnou kontrolu, usnadňuje porozumění zadávanému výrazu a odstranění chyb. Kromě bublinové nápovědy pak dalším prvkem uživatelského rozhraní je funkce našeptávače (autokompletace) pracující s databází odborných matematických výrazů.

Vyhledávání může být zpřesňováno pomocí fasetového rozhraní zadávání dotazů, umožňujícího uživateli postupně ladit dotaz pro co nejlepší ukojení jeho informačních potřeb. Seznam výsledků je pro přehlednost obohacen o zobrazení úryvků z vyhledaných dokumentů s barevným zvýrazněním podčástí relevantních dotazů.

Zájemci o hlubší vhled do práce vyhledávacího systému mohou využít ladícího režimu, který u seznamu výsledků poskytuje podrobné informace o způsobu výpočtu konečného skóre (ranking) každého výsledku.

## **Similarity Search for Mathematics: The Winning Strategy of the NTCIR-11 Math-2 Task**

Martin Líška, Petr Sojka, Michal Ružička

Poster popisuje a sumarizuje skúsenosti tímu MIRMU z Masarykovej univerzity s vyhľadávaním matematiky, ktoré sa uskutočnilo v rámci NTCIR-11 Math-2 vyhľadávacej úlohy. NTCIR (NII Testbeds and Community for Information access Research) je tradičná súťaž v oblasti získavania informácií (IR), ktorá sa zameriava na porovnávanie vyhľadávacích techník vyvíjaných pre rôzne domény IR a koná sa v Japonsku. Náš prístup k vyhľadávaniu matematického obsahu je podobnostné vyhľadávanie založené na kanonikalizácii MathML a druhej generácii škálovateľného plnotextového systému Math Indexer and Searcher (MIaS) využívajúceho overené vyhľadávacie techniky. Schopnosti systému MIaS ako napr. tokenizácia matematických výrazov, unifikácia, kombinovanie matematických a textových výrazov v dotaze boli overené poslaním viacerých vyhľadávacích behov, z ktorých každý bol založený na jednom zo štyroch poskytnutých notačných formátov, kde konečné výsledky vyhľadávania vznikli zlúčením výsledkov z viacerých poddotazov. Dôležitosť kanonikalizácie MathML zápisov matematiky bola overená na jednom zo zaslaných behov, ktorý bol založený výlučne na prezentačnom MathML. Posledný rok sme pracovali na zlepšení kanonikalizácie čo sa ukázalo na zvýšení presnosti výsledkov Pmath behu oproti poslednému vyhodnoteniu na NTCIR-10. Taktiež experimenty v kombinovaní textových slov s matematickými výrazmi ukázali, že by sa malo viac pozornosti venovať budovaniu konečných dotazov z užívateľského vstupu, čo môže mať nezanedbateľný vplyv na presnosť výsledkov vyhľadávania.

## Annotation Game for Textual Entailment Evaluation

Zuzana Nevěřilová

Textual entailment in natural language processing (NLP) is a directional relation between text fragments: text T entails hypothesis H if H follows from T. Textual entailment has more relaxed definition than pure logical entailment: "T entails H" (T ? H) if, typically, a human reading T would infer that H is most likely true.

Recognizing textual entailment (RTE) has been proposed as a generic NLP task in 2004. RTE systems that decide if T ? H is true or false have to be evaluated against manually annotated collection of pairs hypothesis-text (H-T pairs). Such collection can be used for training or testing a RTE application only if it is large enough.

We present a game that collects H-T pairs. It follows a detective story narrative pattern: a dialogue between the brilliant detective and his assistant. In the game the detective (human player) provides a short story. The assistant (the application) proposes hypotheses the detective judges true, false or non-sense.

The textual entailment generation is a rule-based process but the output hypotheses are ranked according to a statistical language model.

The game is intended to collect data in the Czech language. However, the ideas can be applied for other languages as well.



## **Diverse Queries and Feature Type Selection for Plagiarism Discovery**

Šimon Suchomel, Jan Kasprzak, Michal Brandejs

A program for helping deterring real-world plagiarism needs to accomplish many tasks. Original documents which served for creation of plagiarism must be retrieved and also suspicious passages according to input document must be highlighted. This poster presents methodology used during PAN2013 competition on uncovering plagiarism.

The source retrieval task is divided into 2 subtasks: Quering and Selecting, during which the software utilizes a given search engine. The retrieved sources must be examined in detail in order to highlight as many plagiarism cases as possible. This process is depicted as Text Alignment. Results of this process are called detections, i.e. passages of source document and suspicious document, which are similar enough to each other, and can serve as a basis for further manual examination for possible plagiarism. The results of the PAN show that this approach is one of the best for a real-life adoption, since it achieved a decent recall with just a fraction of used queries. Such approach is applicable for detection of suspicious texts, which may contain plagiarism, that can then be selected for further investigation.

## Creating 70+ Billion Word Corpus in the Sketch Engine

Lucia Kocincová, Miloš Jakubíček, Pavel Rychlý

We describe the process of creation of a 70 billion word text corpus of English. We used an existing language resource, the ClueWeb09 dataset [1], as source for the corpus data. Processing such a vast amount of data presented several challenges, mainly associated with pre-processing (boilerplate cleaning, text de-duplication) and post-processing (indexing for efficient corpus querying using the CQL – Corpus Query Language) steps. JusText [2], a boilerplate cleaning tool, was used to remove metadata and low quality text. A deduplication tool Onion [2] was applied to full (document-level) duplicates and also on the level of near-duplicate paragraphs of text. The data was morphologically annotated by part-of-speech tags using TreeTagger [3] to add linguistics features important for language analysis such as an automatic thesaurus and word sketches (one-page, automatic, corpus-derived summaries of a word's grammatical and collocational behaviour). An effective parallelization of the corpus indexation procedure was employed within the Manatee corpus manager [4] (a part of the Sketch Engine corpus management system). The resulting corpus has become a valuable resource for English language analysis.

### References

- [1] The Lemur Project at Carnegie Mellon University, <http://lemurproject.org/clueweb09/>
- [2] Jan Pomikálek. Removing Boilerplate and Duplicate Content from Web Corpora. Phd thesis, Masaryk University. 2011.
- [3] Helmut Schmid. Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK. 1994.
- [4] Kilgarriff Adam, et al. The Sketch Engine: Ten Years On. In *Lexicography* (2014): 1-30.

## **Partial Grammar Checking for Czech Using the SET Parser**

Vít Suchomel, Vojtěch Kovář

Checking people's writing for correctness is one of the prominent language technology applications. In the Czech language, punctuation errors and mistakes in subject-predicate agreement belong to the most severe and most frequent errors people make, as there are complex and non-intuitive rules for both of these phenomena. At the same time, they include numerous syntactic, semantic and pragmatic aspects which makes them very difficult to be formalized for automatic checking. In this paper, we present an automatic method for fixing errors in commas and subject-predicate agreement, using pattern-matching rule-based syntactic analysis provided by the SET parsing system. We explain the method and present an evaluation of the overall accuracy.

## **Formal Methods in High-level Motion Planning for Robotic Systems**

Mária Svoreňová, Ivana Černá, Calin Belta

Robotics is a vast and mature area that concentrates on design and control of reliable computer-based systems. Motion planning for robotic systems aims to produce a continuous motion of a robot that satisfies given movement constraints and/or optimizes some aspect of the movement. A typical motion planning problem is to design a motion that connects a start configuration and a goal configuration of the robot, while avoiding collision with obstacles.

As many robotic systems are safety-critical systems whose failure could result in loss of resources, significant property damage, damage to environment or even loss of life, there is a strong need for rigorous techniques in design of such systems. Formal methods is a term used to refer to techniques from theoretical computer science such as graph or game theory, automata theory, formal verification such as model checking, temporal logics, and others. The deployment of these well-established theories in motion planning allows to generate motion with provable mathematical guarantees of desired properties.

In this work, we first introduce the main concept of applying formal methods in motion planning for robotic systems. We mention several interesting areas of motion planning, where these techniques are being successfully deployed. Finally, we present our results in this area that has been published recently.

## **Management by Competencies**

Petr Štěpánek, Leonard Wallezky

Service oriented companies must very often solve an important, life-based problem — how to motivate their employees to stay and work for the company on one side and how to handle with neverending process of development of new technologies, processes and knowledge on the other side. To hold those two elementary problems in synergy became a serious problem especially identified in IT companies.

As the reaction on this situation the new way of the company management was developed by Plamínek and Fišer (2005) in Czech Republic. It is called Management by Competencies (MbC).

In our poster we want to show the basic principles, taught in the presence on Faculty of Informatics in Service Science, Management and Engineering study program at Masaryk University. This management method, developed for the companies with high added value of the human work, can help those companies, who are faced to the described problems.

Moreover, application of MbC is not limited to the business, but can be applied everywhere, where is necessary to work in teams of experts.

## Device-independent randomness extraction for arbitrarily weak min-entropy source

Jan Bouda, Marcin Pawłowski, Matej Pivoluska, Martin Plesch

High quality randomness is a very useful resource in many computation and cryptographic tasks. In fact it has been shown that many protocols (including quantum ones) vitally require perfect randomness for their security.

Unfortunately, even though we cannot fully predict certain processes it is very difficult to argue that they produce *perfect randomness* – independent and unbiased bits. The problem of imperfect randomness has a long history in classical computer science and long line of research was devoted to randomness extraction – algorithms to transform imperfect sources of randomness into (close to) perfect ones.

The drawback of randomness extractors are twofold. Firstly, extractors typically require at least two independent sources of (imperfect) randomness. Worse still, even imperfect randomness of classical processes has to be assumed, because in principle classical physics is deterministic. Quantum physics, with its intrinsic randomness allows us, in theory, to drop the second assumption. Preparation of a pure state and measurement in its complementary basis will yield a perfectly random result. In practice, however, we are replacing the assumption of randomness by yet another assumption – perfect control of quantum devices. This assumption is also very problematic, as we have learned in case of quantum key distribution.

Luckily enough, thanks to Bell-type experiments, it is possible to certify by classical procedures that quantum processes by are being observed and therefore intrinsic randomness is being produced. This is the basic idea behind device independent randomness extraction. Effectively, we are exchanging the assumption of independent randomness of the second source by a much weaker assumption – validity of quantum mechanics. Alternatively, one can view device-independent randomness extraction as quantum protocol of extracting randomness from a single weak source – a task that is classically impossible.

In this paper we use min-entropy block source with  $n$ -bit blocks of output with guaranteed min-entropy  $H_\infty$ . A block source is called an  $(n, k)$  source if it can be modeled as a sequence of  $n$ -bit random variables  $X_1, X_2, \dots$ , such that  $\forall x_1, \dots, x_{i-1} \in \{0, 1\}^n, \forall e \in \mathcal{I}(E), H_\infty(X_i | X_{i-1} = x_{i-1}, \dots, X_1 = x_1, E = e) \geq k$ .

Therefore, each new block has high min-entropy, even conditioned on the previous ones and any information of the adversary  $E$ . This is a generalization of Santha-Vazirani sources, which can be viewed as block sources with  $n = 1$ .

Note that the task of classically transforming a single block source into a fully random bit is known to be impossible. Furthermore, it is impossible to turn a block source with  $n > 1$  into Santha-Vazirani source, therefore we cannot use existing randomness extraction protocols.

## Surface-Based Visualization of Human Face Variation

I. Chalás, Z. Ferková, P. Urbanová, B. Kozlíková, Z. Kotulanová, J. Sochor

Many current research and clinical disciplines exploit the latest advances and continuous improvements in hardware, computer graphics and vision. By capturing photo-realistic 3D representations of the human body and processing them using various algorithms and techniques, researchers and clinicians are able to perform various computer-aided comparisons, surgery planning or diagnostic assessments automatically. We present our approach to the computation and visualization of morphological variations when human faces are the areas of interest.

The presented algorithm takes as the input a dataset of high resolution 3D scans of human faces. We utilize the novel 3D Virtual Model Database of Human Faces (established at authors' home institution). The whole process can be divided into two parts - the alignment and the comparison of human faces. We introduce a new approach to the creation of an average face derived from the input set of faces. From the final average face and the input set we can determine the variation of the set by calculating the Hausdorff Distance between the average face and each face of the input set. The resulting values are then used to create the color map which is applied to the final average face and forms one of the main results of the algorithm.

The accuracy of the resulting map is highly dependent on the number of iterations of the algorithm and the resolution of the input meshes. The visualization of the final variance using this representation is comprehensible. It highlights the parts where the difference is significant.

These improvements may take part in a beneficial research tool utilized in developmental, cross-population or clinical studies where 3D face records are being examined. They have the potential to be implemented into routine diagnostics in order to facilitate pre- or post-operation assessments or communications among clinicians or/and laymen.

The algorithm was integrated to the novel Forensic 3D Facial Identification Software, named FIDENTIS (being developed at authors' home institution). This tool serves for the analysis of human faces enabling forensic anthropologists to perform a variety of research tasks.

## Bit-Precise LTL Model Checking

Petr Bauch

Complete verification of unmodified code is a challenging task, well-motivated by the costs of software debugging. In this work we rise to the challenge by proposing a model checking method that operates on unmodified parallel programs, specifically accepting LLVM bitcode as input. Apart from being complete, our method proves correctness of a program w.r.t. a temporal specification and is sound w.r.t. arithmetic overflows of integer variables. To overcome the limitations of classical model checking: state space explosion, state matching, etc. we further propound to reduce the model checking problem to a specific instance of the non-termination checking and lift the recently proposed property directed reachability to compute approximations of recurrent sets.