

INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

**Anais do VI Workshop de Teses, Dissertações
e Trabalhos de Iniciação Científica em
Andamento do IC-UNICAMP**

Jefferson L. M. da Silveira *Ana C. C. Rézio*

Ariadne M. B. R. Carvalho *Fábio A. Faria*

Bruno. C. M. Vilar *Gustavo P. Alkimin*

Roberto Pereira *Tiago R. C. Falcão*

Technical Report - IC-11-13 - Relatório Técnico

June - 2011 - Junho

The contents of this report are the sole responsibility of the authors.
O conteúdo do presente relatório é de única responsabilidade dos autores.

Apresentação

Este relatório técnico contém os resumos de 26 trabalhos apresentados no VI Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica, do Instituto de Computação da UNICAMP (WTD-IC-UNICAMP 2011). O Workshop ocorreu no dia 11 de maio de 2011. Na ocasião tivemos apresentações orais, feitas por alunos de mestrado e doutorado, e um feira de pôsters. Aos alunos de pós-graduação foi dada a possibilidade de escolher a forma de apresentação (oral, pôster, ou ambas), e aos alunos de iniciação científica coube a apresentação na forma de pôster. A publicação dos resumos sob a forma de um relatório técnico tem por objetivo divulgar os trabalhos em andamento e registrar, de forma sucinta, o estado da arte da pesquisa do IC no ano de 2011. Como coordenadora do VI Workshop, agradeço imensamente o trabalho dos alunos do programa de pós-graduação do IC que efetivamente organizaram o evento e que são co-editores deste relatório – Ana, Bruno, Fábio, Gustavo, Jefferson, Roberto e Tiago. Agradeço também aos alunos que participaram do evento, em particular aqueles que se dispuseram a apresentar seus trabalhos, seja oralmente ou na forma de pôsters, bem como aos seus orientadores que os incentivaram a fazê-lo. Finalmente, agradeço à secretaria e à coordenação de pós-graduação do IC, em especial ao professor Alexandre Falcão, pelo incentivo, apoio e patrocínio ao evento.

Ariadne Maria Brito Rizzoni Carvalho

Coordenadora do VI WTD

Professora do Instituto de Computação - UNICAMP

Sumário

1	E-Learning Environment with Multimodal Interaction: A Propose to Improve the Usability, Accessibility and Learnability of e-Learning Environments	4
2	Leilões e Algoritmos	6
3	Fitting 3D Deformable Biological Models to Microscope Images	8
4	Implementação eficiente de algoritmos criptográficos em arquiteturas modernas	10
5	Security Data Parametrization and Visualization	12
6	Multiscale Parameter Search (MSPS)	14
7	Um Oráculo para Teste de Robustez Baseado em Algoritmos de Alinhamento de Sequências	16
8	Gerenciamento de dados de sensores para Sistemas de informação de Biodiversidade	18
9	Managing Data Provenance for Biodiversity Information	21
10	Paralelização de Simuladores Descritos em <i>SystemC</i>	23
11	Arquitetura para Suporte a Aplicações Ubíquas que Viabilizam a Criação de um Ambiente de Aprendizado Ativo em Sala de Aula	25
12	Geração Automática de Teste Baseado em Modelo	27
13	Segmentação interativa de objetos em imagens e vídeos baseada em grafos e modelos nebulosos de conhecimento de conteúdo	29
14	Método Ágil aplicado ao Desenvolvimento de Software Confiável baseado em Componentes	31
15	Análise Comportamental de Código Malicioso Através da Monitoração de Chamadas de Sistema e Tráfego de Rede	33
16	Integrating and Enriching Folksonomies and Ontologies	35
17	Explicitação de Esquema Orientada a Contexto para Promover Interoperabilidade Semântica	37
18	Complexidade de Construção de Árvores PQR	39

19	Abordagem para Interoperabilidade Semântica em Bases de Dados Multi-Representação	41
20	Algoritmos Aproximados para Problemas de Empacotamento em Faixa com Restrições de Descarregamento	43
21	Shadow-driven Annotation	45
22	Empacotamento Tridimensional e Programação por Restrições	47
23	Simulação de multidões e planejamento probabilístico para otimização de intersecções controladas por semáforo	49
24	Visualization Techniques for Content-Based Image Retrieval	51
25	Rearranjos em Genomas	54
26	Aplicação de Técnicas de Visão Computacional e Aprendizado de Máquina para a Detecção de Exsudatos Duros em Imagens de Fundo de Olho	56
27	Programação do WTD	58

1 E-Learning Environment with Multimodal Interaction: A Propose to Improve the Usability, Accessibility and Learnability of e-Learning Environments

Authors: André Constantino da Silva, Heloísa Vieira da Rocha

The Human-Computer Interaction is challenging the replacement of the mouse and actual interfaces for interfaces that works with natural interaction, non-tactile interfaces, speech recognition, facial and movement recognition and gestures [2] [3], embracing the support of different interaction styles. Faced by the many hardware that users have to interact with the application and the variety of physics and social context we need to study the use of different interactions styles and their simultaneous use. And good access is not yet sufficient, it is necessary to take advantages from the interaction styles.

In this context, we argue as human-computer interface specialists how the software can be used in different devices by the diversity of users and physical and social contexts so that they can get the better of the interaction style?

It is the main problem, but several others derive: Is it possible to improve an application to be used by many interaction styles and get advantages? Which kind of modifications is necessary to get the best of the chose interaction style? How to guaranties that the usability will not be affected? After the modifications, the application has a better accessibility? Allowing many interaction styles, do we have a new kind of application? How to differentiated the applications that have these characteristics from the other that does not have?

We argue it is possible to improve an application so that the users can get the best of the devices, not thinking directly about the devices, but supporting all interested interaction styles. One way to solve the research problem is allow users to interact with the environment in different modes, e.g. visual, touch, voice or gesture, in a good way, in our case, a good usability application.

In special, we chose the e-Learning domain to study. The special interest in this domain is the necessity to improve the e-Learning environments, like TelEduc [5] and Ae [1], to better attend the teaching and learning activities in the variety of learning context. So we ask: can the support for multiple interaction styles improve the environment so that it can support better any context learning? We call learnability this property to support any context learning.

The adopted method consist into observer many users (teachers and students) using an e-Learning Environment with many interaction styles, collecting advantages and disadvantages during the use. Not one interaction style will be studied per time; we want study various interaction styles being used to complete some tasks in the environment. But firstly, we need implement an e-Learning Environment so that users can interact with many interaction styles. This is not a trivial task, because the literature does not have sufficient information to how design applications for many interaction styles, and there is a lack of technologies to support them.

The use of Multimodal Interaction shows a good and viable solution to possibility the use of the e-Learning environment in this diversity of users, devices, contexts and interaction styles. Multimodal Interaction is proposed to turn the human-computer interaction more

natural, i.e., more close to the human-human interaction. The main benefices are the increase of application's usability, accessibility, flexibility and convenience [4].

So, this work aims to study the impact of Multimodal Interaction in the e-Learning Environments. To reach this goal, we propose to implement the Multimodal Interaction concepts in the Ae e-Learning environment [1], and use this new kind of environment to research advantages and disadvantages of multimodal interfaces in learning. We planned study the interaction using pen and finger in two tools of the Ae e-Learning environment: the Weblog and the Whiteboard. Both tools are using to construct content and are selected based on our premises that they are good choices to study the multimodality.

This work is related with some other works realized in the IC/UNICAMP to study the use and impact of the use of new technologies in the education, and we acknowledge the CAPES financial support.

The main expected contribution is to know more about the implementation and use of multimodal interaction, more precisely, using pen, touch and gesture. In the educational context, if we have an e-Learning environment with multimodal interaction, can we embrace more learning contexts? If this question is yes, so we have another property for environments: the learnability.

Referências

- [1] Ae Environment. *<http://tidia-ae.iv.org.br/>*. 2011.
- [2] B. Hayes. Gartners Seven IT Grand Challenges. *Communications of ACM*, 51(7), July 2008.
- [3] L. Kugler. Goodbye, Computer Mouse. *Communications of ACM*, 51(9), September 2008.
- [4] D. Lalanne, L. Nigay, P. Palanque, P. Robinson, J. Vanderdonckt, J. Ladry. Fusion engines for multimodal input: a survey. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pages 153–160, Cambridge, Massachusetts, USA, 2009.
- [5] TelEduc Environment. *<http://www.teleduc.org.br>*. 2011.

2 Leilões e Algoritmos

Autores: Carlos Eduardo de Andrade e Flávio Keidi Miyazawa

Os leilões tem sido, a vários séculos, a maneira mais generalizada de alocação de itens e recursos quando os agentes envolvidos na transação não têm certeza do valor dos itens envolvidos. Um leilão permite que exista uma disputa entre os interessados que, através de um protocolo de negociação, são levados a revelar o quão importante é um item que negocia. É latente a expansão da utilização de leilões na última década dada a expansão da internet que é usada como principal meio de comunicação para a realização das atividades.

Os leilões podem ser vistos como uma classe de problemas de alocação clássicos mas, diferentemente destes últimos, os leilões possuem características que os definem em um escopo diferente. A primeira é que o conjunto de agentes envolvidos interagem de *maneira estratégica* e são *racionais* no sentido que possuem como objetivo maximizar seu *payoff*. Segundo, cada agente mantém um conjunto de informações privadas que apenas ele conhece deterministicamente. Isto torna o sistema descentralizado onde cada agente possui informação incompleta. Terceiro, cada agente pode escolher uma estratégia que está disponível para ele e deve ter “inteligência suficiente” para determinar estratégias de melhor resposta. Estas três características fazem dos leilões problemas de teoria dos jogos, onde temos que resolver um problema de decisão ou otimização com informação ou especificação incompleta (vide Krishna [4] para mais detalhes).

Entre os diversos modelos de leilões, destacamos os *Leilões Combinatoriais* onde os licitantes (compradores ou vendedores, dependendo do contexto) podem dar seus lances em combinações de itens, chamados “pacotes”, ao invés de apenas itens individuais. Estes tem sido usados recentemente para tratar problemas como transporte de cargas, alocação de rotas, alocação de recursos naturais (espectro de rádio em comunicações sem fio, sítios de prospecção de óleo e gás natural), pregões e contratos.

Definição: Seja $N = \{1, 2, \dots, n\}$ o conjunto de licitantes e $M = \{1, 2, \dots, m\}$ o conjunto de itens. Cada licitante i tem uma função $v_i : 2^M \rightarrow \mathbb{R}$ que mede seu grau de preferência ao receber um determinado conjunto de itens $S_i \in 2^M$, uma função $b_i : 2^M \rightarrow \mathbb{R}$ que é o valor oferecido pelo licitante para este conjunto, e uma função de utilidade $u_i(S) = v_i(S_i) - p_i$, onde p_i é o preço pago pelo licitante i (u_i é o *payoff* de i). Uma *alocação* de itens $S = (S_1, \dots, S_n)$ para os licitantes é tal que $S_i \cap S_j = \emptyset$ para $i \neq j$. O *bem-estar social* obtido por esta alocação é $u(S) = \sum_i u_i(S_i)$. Uma *alocação socialmente eficiente* é uma alocação que maximiza o bem-estar social entre todos licitantes. A *receita* do leiloeiro é $r(S) = \sum_i p_i$, e S é dita ser uma *alocação ótima* se maximiza esta receita. Note que ambos termos “eficiente” e “ótimo” são definidos em contexto econômico, em contraste com seus significados usuais em computação.

Note que os leilões combinatoriais potencialmente apresentam instâncias muito grandes. Se um determinado licitante pode dar um lance em qualquer subconjunto de itens, são possíveis 2^m lances por licitante, que é um número muito elevado mesmo para m de tamanho modesto. Assim, temos que criar soluções econômicas e computacionais para mitigar esse problema. Uma delas é definir linguagens de lances compactas. Nisan [2, capítulo 9]

apresenta um compêndio dedicado a linguagens de lances.

Outro problema é que a determinação dos vencedores do leilão é um problema \mathcal{NP} -difícil e é inaproximável para o caso geral [6]. Este problema é considerado o elemento central no projeto de leilões, pois a maneira como se determina os ganhadores e seus pagamentos influencia diretamente nos lances efetuados pelos licitantes. Em geral, buscamos um leilão *incentivo-compatível* onde, por definição, os licitantes sempre oferecerão os valores reais que atribuem aos itens. Este contexto é necessário para a maximização do bem-estar social e da receita do leiloeiro (quando não conhecemos as distribuições de probabilidade que regem os lances de cada licitante). Infelizmente, a utilização de heurísticas e as aproximações para o caso geral destroem a estrutura de incentivo-compatibilidade requerida [1].

Assim, grande esforço tem sido feito nos últimos anos para encontrar cenários e algoritmos que permitam resolver o problema de determinação dos ganhadores e pagamentos e que sejam incentivo-compatíveis. Alguns cenários, como leilões multiunidade [3] e de pacote único [5], apresentam resultados positivos. Mas ainda pouco se conhece do comportamento de leilões com outros tipos de restrições.

Questões em Aberto: *Quais são as características que um leilão combinatorial possui e que garantam a existência de mecanismos de alocação incentivo-compatíveis em termos de eficiência econômica e computacional? Quais são os limitantes e aproximações possíveis?*

Em resumo, embora os leilões sejam estudados há mais de meio século, o estudo das propriedades algorítmicas aplicadas a eles é uma área relativamente recente e que carece de muito estudo. A principal vertente é o estudo da incentivo-compatibilidade e da maximização da receita, já que ambas são conflitantes entre si e com requerimentos algorítmicos.

Referências

- [1] D. Buchfuhrer, S. Dughmi, H. Fu, R. Kleinberg, Y. Singer, E. Mossel, C. Umans, C. Papadimitriou e M. Schapira. Inapproximability for VCG-Based Combinatorial Auctions. *Proceedings of the 21th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA'10*, páginas 518–536, 2010.
- [2] P. Cramton, Y. Shoham e R. Steinberg. *Combinatorial Auctions*. MIT Press, 2006.
- [3] S. Dobzinski, R. Lavi e N. Nisan. Multi-unit auctions with budget limits. *Proceedings of FOCS'2008*. IEEE Computer Society, páginas 260–269, 2008.
- [4] V. Krishna. *Auction Theory*. Academic Press, 2a edição, 2010.
- [5] D. Lehmann, L. I. O'Callaghan e Y. Shoham. Truth revelation in approximately efficient combinatorial auctions. *Journal of ACM*, 49:577–602, 2002.
- [6] T. Sandholm. Algorithm for optimal winner determination in combinatorial auctions. *Artificial Intelligence*, 135(1–2):1–54, 2002.

3 Fitting 3D Deformable Biological Models to Microscope Images

Authors: Danillo Roberto Pereira and Jorge Stolfi

The popularity of digital microscopes created an acute shortage of manpower to analyze and classify their these images. In clinical microscopy examinations, for example, the analysis of a single image that is produced in a matter of seconds may demand hours of visual inspection. We consider here a common type of analysis which is the matching a given image of certain organisms to an abstract model of the same. This thesis aims to produce tools to automate this task.

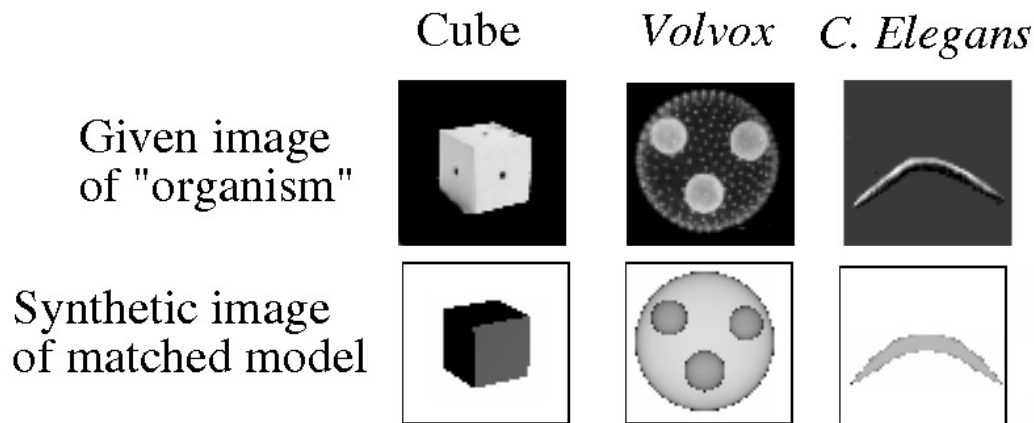
We are mainly concerned with biological structures smaller than ≈ 1 mm, imaged through a microscope of some sort (to simplify the discussion we will refer to such structures as *organisms*, even though our method can be applied also to isolated cells of multicellular organisms, organs, organelles, etc). Each kind of organism, at a specific stage of life, usually has recognizable morphology.

The organism matching task is made harder by the translucent nature of most microscopic biological structures, and by the complexities of the microscope imaging process — which may include very short depth-of-focus, interference and diffraction artifacts, and the common occurrence of dust and other debris over and around the organism. A major difficulty in this problem is the fact that the same organism can produce very different images, depending on its orientation in space. Futhermore, many organisms of interest have gelatinous and/or elastic consistency, and may suffer considerable deformation. Yet, with some experience, a human observer can identify microscopic images of the organisms of various kinds, and align them with their three-dimensional schematic diagrams.

Multiscale alignment: In real situations, the position and deformation of the model may vary over a wide region of the space \mathcal{D} of parameters, whose radius ρ is much larger than the required accuracy δ . A “brute force” solution would enumerate a grid of points in the region \mathcal{D} with step δ . However, the number of such points is the order of $(\rho/\delta)^n$, where n is the number of parameters. For typical values of ρ , δ and n , this number is astronomical.

The exponential cost barrier can be overcome by *multiscale matching* techniques. Multiscale techniques use two or more versions $A^{(0)}, A^{(1)}, \dots, A^{(m)}$ of the entity A involved — in our case, the input digital image A — each with a different level of detail. The problem at hand is solved first for the coarsest scale version $A^{(m)}$ yielding a tentative set of parameters, $p^{(m)}$. This solution is then used as an initial guess to solve the problem for the next scale $A^{(m-1)}$, where there is a bit more detail [1, 3, 2] yielding an improved parameter vector $p^{(m-1)}$. This process is repeated until reaching the original scale $A^{(0)}$. Since only small adjustments are considered at each stage, the total running time is greatly reduced in comparison with the brute-force approach. Since the best match at a coarse scale may not correspond to the best match at scale 0, it is necessary to keep several candidates matches at the coarsest scale, and reduce the size of this set as the matching progresses to the finer scales.

Preliminary Tests: For these tests we used a few simple “organisms”, comprising one geometric object (a cube) and simple real micro-organisms (*Volvox* sp., and *Caenorhabditis elegans*).



Conclusions and future improvements: The tests show that our algorithm can find a deformable organism model in a given image. Moreover, the method can be generalized to other models and other applications. The current implementation can handle about 10 parameters. In the future, we plan to use non-linear optimization methods to allow us to handle models with dozens of parameters.

Acknowledgment. This research is supported by FAPESP (grant 2007/52015-0 and 2009/13744-2) and CAPES.

Referências

- [1] H. C. da Gama Leitão and J. Stolfi. A multiscale method for the reassembly of two-dimensional fragmented objects. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1239–1251, 2002.
- [2] C. E. A. Zampieri. *Recuperação de imagens multiescala intervalar*. Master thesis, UNICAMP, BRAZIL., 2010.
- [3] D. R. Pereira and J. Stolfi. Fitting 3d deformable models to microscope images. *L&NLM - Learning and Nonlinear Models*, (1), 2011.

4 Implementação eficiente de algoritmos criptográficos em arquiteturas modernas

Autores: Diego Aranha¹, Julio López

A criptografia baseada em emparelhamentos possibilitou a flexibilização das primitivas criptográficas conhecidas e ampliou de forma considerável os cenários de aplicação de criptografia assimétrica. Entre os protocolos baseados em problemas sobre grupos bilineares, destacam-se: acordo de chaves eficientes para múltiplas entidades [1], assinaturas curtas [2] e paradigmas alternativos de certificação implícita [3].

Apesar das propriedades inovadoras, o desempenho de sistemas criptográficos baseados em emparelhamentos ainda representa um obstáculo. Tipicamente, o cálculo de um emparelhamento bilinear ainda é significativamente mais caro do que uma execução de um protocolo variante do RSA [4], e uma ordem de magnitude menos eficiente que uma multiplicação de ponto em curvas elípticas [5]. Isto é natural, visto que os métodos mais estabelecidos de criptografia assimétrica puderam receber maior esforço de pesquisa, produzindo algoritmos cada vez mais eficientes.

Este projeto tem como finalidade desenvolver algoritmos eficientes (seqüenciais e paralelos) e implementações em *software* otimizadas para criptografia baseada em emparelhamentos. A criptografia baseada em emparelhamentos bilineares emprega aritmética em múltiplos níveis [5]: algoritmos para o cálculo do emparelhamento propriamente dito [6, 7], aritmética na curva elíptica onde o emparelhamento bilinear encontra-se definido, aritmética no corpo finito onde a curva elíptica está definida e aritmética nas extensões deste corpo finito.

O desenvolvimento destes novos algoritmos, além de produzir resultados teóricos relevantes, também deve acompanhar as tendências tecnológicas atuais para que sua implementação considere os recursos disponíveis na máquina para obtenção de desempenho. Para isso, paralelismo em nível de tarefas e em nível de dados são extensamente utilizados, incluindo a aplicação de multiprocessamento e conjuntos de instruções vetoriais. Como principais resultados já obtidos, pode-se citar:

1. **D. F. Aranha**, R. Dahab, L. B. Oliveira, J. López. *Efficient implementation of elliptic curve cryptography in wireless sensors*. Advances in Mathematics of Communications, Volume 4, Issue 2, pp. 169-187, 2010.
2. L. B. Oliveira, **D. F. Aranha**, C. P. L. Gouvêa, M. Scott, D. F. Câmara, J. López and R. Dahab. Computer Communications, Volume 34, Issue 3, pp. 485-493, 2011.
3. **D. F. Aranha**, J. López, D. Hankerson. *High-speed parallel software implementation of η_T pairing*. In: CT-RSA 2010, pp. 89-105, 2010.
4. **D. F. Aranha**, J. López, D. Hankerson. *Efficient Software Implementation of Binary Field Arithmetic Using Vector Instruction Sets*. In: LATINCRYPT 2010, pp. 144-161, 2010.

¹Financiado por FAPESP, processo 2007/06950-0, e CAPES, processo 5551-09-9.

5. **D. F. Aranha**, K. Karabina, P. Longa, C. H. Gebothys, J. López. *Faster Explicit Formulas for Computing Pairings over Ordinary Curves*. In: EUROCRYPT 2011, To appear.

Os dois primeiros aprimoram o estado da arte de implementações de criptografia em redes de sensores sem fio, ampliando sua viabilidade e eficiência. A aritmética eficiente no corpo permitiu o cálculo de uma multiplicação de ponto 61% mais rápida que a melhor implementação de curvas elípticas sobre corpos binários e 57% mais rápida que a melhor implementação para o caso primo, considerando o mesmo nível de segurança, e uma aceleração de 22% no cálculo do emparelhamento η_T sobre curvas supersingulares binárias. Os dois trabalhos seguintes investigam implementações do Algoritmo de Miller para o cálculo de emparelhamentos, explorando conjuntos de instruções vetoriais e multiprocessamento. Em particular, o estado da arte de implementações de emparelhamentos sobre corpos binários em uma arquitetura *Intel Core* de 64 bits foi aprimorado em 24%, 28%, 44% e 66% com o emprego de 1, 2, 4 e 8 processadores. O último trabalho aprimora o estado da arte anterior para o cálculo de emparelhamentos sobre corpos primos em até 34% com uso de um conjunto de parâmetros e novas fórmulas mais eficientes para a aritmética exigida.

Referências

- [1] A. Joux. A one round protocol for Tripartite Diffie-Hellman. *Journal of Cryptology*, 17(4):263–276, 2004.
- [2] D. Boneh, B. Lynn, and H. Shacham. Short signatures from the Weil pairing. *Lecture Notes in Computer Science*, 2248, 2001.
- [3] S. S. Al-Riyami and K. G. Paterson. Certificateless public key cryptography. In *ASIACRYPT*, pages 452–473, 2003.
- [4] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 26(1):96–99, 1983.
- [5] M. Scott. Implementing cryptographic pairings. Tech Report, 2009. <ftp://ftp.computing.dcu.ie/pub/resources/crypto/pairings.pdf>.
- [6] P. S. L. M. Barreto, H. Y. Kim, B. Lynn, and M. Scott. Efficient algorithms for pairing-based cryptosystems. In *CRYPTO '02*, pages 354–368, 2002. Springer.
- [7] P. S. L. M. Barreto, S. Galbraith, C. Ó hÉigearthaigh, and M. Scott. Efficient Pairing Computation on Supersingular Abelian Varieties. *Design, Codes and Cryptography*, 42(3):239–271, 2007.

5 Security Data Parametrization and Visualization

Authors: Gabriel Dieterich Cavalcante, Sebastien Tricaud and Paulo Lício de Geus

Nowadays, a security administrator is responsible for the analysis of large amounts of data that represent activities on the network. System logs and network traffic can provide useful information to describe what is happening. However, a single service can produce a myriad of lines of log a day. Combined with the complicated topologies of private networks, this really represents an overwhelming volume of data.

Computer forensics involves the preservation, identification, extraction, and documentation of computer evidence stored as data or magnetically encoded information [1]. Normally, computer evidences are transparently created by the computer's operating system, as instructed by its operator through service configuration files. On the other hand, some useful information may be hidden from inspection, and some special forensic tools and techniques are required.

Complicating the whole scenario, computer tools has particularities in the generation of logs, so analyst need know such differences and correlate many kinds of data.

Visualization techniques can help forensic specialists direct their efforts to suspicious events, processes, or hosts, assisting the data interpretation process. Furthermore, one can draw from the visualization world how to approach the problem. In Computer Security, it is important to be able to extract useful information from sparsely clustered events without losing sight of the big picture, i.e. doing this with all data available.

Parallel coordinates[3] is a well recognized method to visualize multivariate data. It uses an n -dimensional representation to reveal correlations in multiple dimensions[4]. Among several techniques available, parallel coordinates have been widely used for visualization of high-dimensional datasets. Parallel coordinates—therefore called *—-coords*—can plot graphs with a huge number of events.

Some rare events in *—-coords* graphs can be revealed when it used interactively. Our idea is use *—-coords* to represent computer security data in a fashion and interactive way, so the analyst can take advantage deeping into data in a more comprehensive manner instead of read thousands of log lines.

From logs to graphs, to transform data into a *—-coords* plot image, the user needs to follow three steps: **data acquisition**, **descriptive language creation**, and **interaction**. Data acquisition is described by name, the forensic professional needs to analyze the largest possible amount of data that can describe something about the scenario in question, so the tools had be prepared to output information about your behavior. The second step is transform logs into a graph description format, for that some scripts is used to parser important information.

To address parametrization problems, we use common description language to create *—-coords* graphs. It is like CVS, but add some useful information about annotated data. This language is called Picviz Description Language—PGDL. The PGDL language has special design to to define *—-coords*, structuring the information, allowing users to define how data will be treated on graph creation.

The PGDL is used by Picviz to transform acquired data into a *—-coords* plot image

to visualize data and discover interesting results quickly. Picviz exports to many figure formats like PNG and SVG, but only static graphs.

Some rare events in `---coords` graphs can be revealed when it used interactively. To address this issue, we developed the Picviz-GUI tool, which complements Picviz—adding interactivity to the visualization of `---coords` graphs. With Picviz-GUI it is possible to shape a graph to reduce visual clutter and to help finding patterns. A set of simple actions (such as filtering, changing line thickness and color, and selections) allows user to highlight the desired information, search through the variables for that subtle data correlation, and thus uncover security issues.

The existing features present on Picviz-GUI could confirm alarms about distributed network scans raised over the Internet by many experts in cyber-threats in July/2010, the GUI depicts some interesting points of new and complex kinds of attack. It has been used data logs from a host in an academic network. In a few steps Picviz-gui has showed the reasons why this type of attack was not averted by existing Intrusion Detection systems. Revealing sophisticated mechanisms of communication and interaction among the botnets.

Picviz can be considered in an early stage, some features are missing specially in process of finding correlated events automatically. However, in the computer forensic area Picviz can be helpful to find rare and unexpected events, as much as understand more complex mechanisms hidden in huge amounts of events.

Referências

- [1] W. Kruse and J. Heiser, *Computer forensics: incident response essentials*. Addison-Wesley, 2008.
- [2] S. Tricaud and P. Saadé, “Applied parallel coordinates for logs and network traffic attack analysis,” *Journal in computer virology*, vol. 6, no. 1, pp. 1–29, 2010.
- [3] A. Inselberg and B. Dimsdale, “Parallel coordinates: a tool for visualizing multi-dimensional geometry,” in *Proceedings of the 1st conference on Visualization’90*. IEEE Computer Society Press, 1990, p. 378.
- [4] J. Yang, W. Peng, M. Ward, and E. Rundensteiner, “Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets,” 2003.

6 Multiscale Parameter Search (MSPS)

Autores: Giovani Chiachia, Alexandre X. Falcão e Anderson Rocha

Muitos problemas práticos de otimização não podem ser formulados analiticamente, assim como não é possível obter informações sobre suas estruturas. Isso acontece, por exemplo, quando o problema é representado por um programa de computador. Nesse caso, a alternativa é guiar a busca pela solução ótima do problema a partir das respostas que este oferece quando soluções candidatas são testadas [1]. Esse tipo de abordagem é conhecida como otimização de *caixa-preta* e métodos capazes de lidar com problemas desse tipo são úteis em função do grande número de cenários em que podem ser aplicados, além de serem fáceis de usar [2, 3].

Cientes de que nenhum algoritmo pode obter a melhor solução em todos os problemas de otimização [2], nós propusemos um método determinístico, ao qual chamamos de Multiscale Parameter Search, para resolver problemas caixa-preta tal como o nosso de registro de imagens. O MSPS pode lidar com problemas não convexos e *não restritos*, definidos em \mathcal{R} , e para os quais não há restrições impostas à solução com exceção dos intervalos de valor em que se deseja obter a solução. A otimização de problemas desse tipo é chamada de *Otimização Global* e a necessidade de resolvê-los emerge de todos os ramos da engenharia e das ciências pura e aplicada [4].

Ao conduzir a busca pela solução em múltiplas escalas de deslocamento, a ideia do MSPS é reduzir o número de execuções da *função-objetivo* (problema a ser otimizado). Além disso, através de uma combinação sinérgica entre as diferentes escalas, o método pode obter um melhor mapeamento sobre o comportamento da função, o que é desejável para lidar com a não convexidade dos problemas. As escalas são definidas de acordo com os intervalos de valores nos quais os parâmetros devem ser otimizados, inerentes do problema, e outras três informações paramétricas: o *número de escalas*, a *taxa de crescimento* das escalas e o *fator de decrescimento do intervalo de amostragem*. Nos experimentos, nós fizemos uso da técnica de *Meta-otimização* para estabelecer de maneira justa os valores ótimos dos parâmetros de controle de cada método levado em consideração [6].

Considere θ como a primeira solução candidata para o problema sendo otimizado e θ_i como os parâmetros do problema, i.e., suas dimensões. A partir desta posição $\theta = (\theta_1, \theta_2, \dots, \theta_n)$, o MSPS procura por deslocamento ótimos Δ^* tal que θ é iterativamente deslocado para novas posições $\theta = \theta + \Delta^*$ até que um critério de otimalidade seja satisfeito ou um número de iterações seja executado. Para que os deslocamentos sejam mais eficazes, o método conduz a busca em múltiplas escalas, perturbando cada parâmetro $i = \{1, 2, \dots, n\}$ em $j = \{1, 2, \dots, m\}$ escalas de deslocamento. A cada iteração, o MSPS avalia o valor de $F(\theta + \Delta)$ para os deslocamentos Δ resultantes das perturbações de cada parâmetro individualmente em todas as escalas e também derivadas da composição das melhores perturbações dentro das escalas e entre elas [5].

Embora nosso foco com o MSPS sempre tenha sido empregá-lo na obtenção de um método de registro de imagens eficiente e eficaz, para colocá-lo frente a métodos de otimização no estado da arte, tais como *Particle Swarm Optimization* (PSO), *Differential Evolution* (DE), e *Simulated Annealing* (SA), nós também conduzimos experimentos con-

siderando uma coleção de 12 problemas *benchmark* amplamente estudados na literatura e difíceis para muitos métodos de otimização [7]. Esses problemas podem ser estendidos para um número arbitrário de dimensões (i.e., parâmetros) e variam entre convexo e não convexo e entre separáveis e não separáveis.

Com os experimentos, dois aspectos gerais do MSPS foram evidenciados. O primeiro é que o MSPS teve o melhor desempenho em seis dos 12 problemas avaliados. Em alguns casos, a vantagem foi considerável. O segundo aspecto é que o MSPS sempre proveu soluções aceitáveis em comparação com os outros métodos. Nos problemas não convexos, o MSPS foi melhor que os outros em todos os casos. Também vale a pena destacar a habilidade do MSPS em fazer o ajuste fino das soluções (observável pela precisão das repostas) e de convergir rapidamente em problemas convexos simples. Esse conjunto de características é exatamente o que nós esperávamos para resolver nosso problema de registro de imagens de faces.

Referências

- [1] Hemker, T.: Derivative free surrogate optimization for mixed-integer nonlinear black box problems in engineering. Ph.D. thesis, Technische Universitat Darmstadt (2008)
- [2] Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* **1**(1), 67–82 (1997)
- [3] Huang, D., Allen, T.T., Notz, W.I., Zheng, N.: Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization* **34**, 441–466 (2006)
- [4] Floudas, C.A., Gounaris, C.E.: A review of recent advances in global optimization. *J. of Global Optimization* **45**, 3–38 (2009)
- [5] Chiachia, G., Falcão, A.X., Rocha, A.: Multiscale Parameter Search (MSPS): A deterministic approach for black-box global optimization. *Journal of Global Optimization* **submitted**, (2011)
- [6] Pedersen, M.E.H.: Tuning & simplifying heuristical optimization. Ph.D. thesis, University of Southampton (2010)
- [7] Yao, X., Liu, Y., Lin, G.: Evolutionary programming made faster. *Evolutionary Computation, IEEE Transactions on* **3**(2), 82–102 (1999)

7 Um Oráculo para Teste de Robustez Baseado em Algoritmos de Alinhamento de Sequências

Autores: Gizelle Sandrini Lemos¹, Eliane Martins (Orientadora).

Testes de robustez determinam se o comportamento do software é aceitável na presença de entradas inválidas ou em condições estressantes [1]. Após sua realização é necessário utilizar um oráculo, referência do comportamento que o software deveria apresentar durante o teste, para avaliar se o resultado condiz ou não com o aceitável. Porém, a determinação do oráculo em testes de robustez não é trivial pois é difícil saber qual será o comportamento do software em situações inesperadas.

O problema de decidir se o resultado da execução do teste foi ou não aceitável é conhecido como *problema do oráculo* [2]. Esse é um problema em aberto pois as soluções usadas atualmente como oráculo para testes de robustez ([3] [4]) não são capazes de afirmar de forma confiável se o comportamento do software foi aceitável.

O objetivo deste trabalho é o desenvolvimento de uma abordagem a ser utilizada como oráculo na avaliação de resultados de teste de robustez, que seja capaz de informar se o comportamento do software é aceitável em relação à propriedades de robustez previamente definidas. As propriedades são verificadas em relação a traços de execução colhidos durante os testes de robustez. Essa verificação ocorre através do alinhamento de traços e propriedades com o uso dos algoritmos de alinhamento de sequências comumente aplicados em bioinformática [5] conforme ilustra a Fig 1.

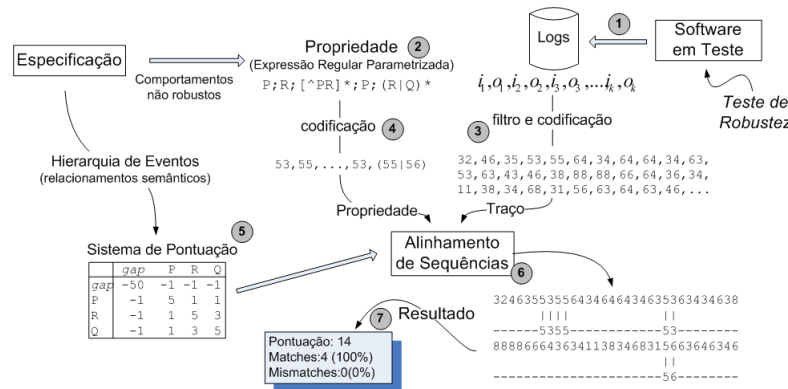


Fig 1: Visão geral da abordagem

Durante os testes de robustez são coletados traços (passo 1) e são definidas propriedades de robustez (passo 2). É necessário convertê-los num formato adequado ao algoritmo de alinhamento (passos 3 e 4). Em seguida, o algoritmo obtém a pontuação do alinhamento entre as sequências (passo 5). O alinhamento é formado de acordo com os valores do sistema de pontuação previamente definido (passo 6). Cada valor do sistema de pontuação é definido com base nos relacionamentos semânticos existentes entre os eventos do software,

¹Trabalho financiado pelo CNPq. Processo 142870/2009-9.

descritos numa hierarquia de eventos idealizada a partir de [6]. Finalmente, são obtidos como resultados a pontuação e o número de *matches*, *mismatches* e *gaps* no alinhamento (passo 7) que nos informam sobre a robustez do traço em relação a propriedade.

A metodologia de trabalho abrange o estudo de áreas de interesse na engenharia de software, mais especificamente, os oráculos de teste. Também foram estudados algoritmos de alinhamento de sequências para o mecanismo de busca da abordagem proposta. A validação dos algoritmos e a análise dos resultados da aplicação da abordagem ocorrerão com a realização de estudos de caso. Além disso, aplicaremos métricas, como falsos positivos e falsos negativos, para avaliar a abrangência do algoritmo em relação aos resultados obtidos durante os experimentos.

Num dos estudos de caso realizados avaliamos diferentes algoritmos de alinhamento de sequências: global, semi-global e local, sendo que o último foi escolhido por ser favorável ao alinhamento de regiões com alta similaridade e pelo fato de que o traço de execução geralmente tem tamanho bem maior do que a propriedade procurada. Em outro estudo de caso avaliamos diferentes sistema de pontuação para o algoritmo de alinhamento [7]. Por fim, foi definida a forma de obtenção dos valores utilizados no sistema de pontuação com a hierarquização dos eventos do software.

Como próximos passos, estudaremos um algoritmo de alinhamento local que usa *affine gaps*. *Affine gaps* minimizam o peso das penalidades *gaps* no alinhamento e permitem que a pontuação de alinhamentos com muitos *gaps* não seja prejudicada.

Referências

- [1] F. Saad-Khorchef, A. Rollet, and R. Castanet. A Framework and a Tool for Robustness Testing of Communicating Software. *ACM Symposium on Applied Computing*. Pages: 1461–1466. 2007.
- [2] W.E. Howden. Theoretical and Empirical Studies of Program Testing. *3rd Int. Conf. on Software Engineering*. 1978.
- [3] H. Hiller, A. Jhumka, and N. Suri. An Approach for Analysing the Propagation of Data Errors in Software. *31st Int. Conf. on Dependable Systems and Networks*. Pages: 161–170. 2001.
- [4] P. Koopman, and J. DeVale. Comparing the Robustness of POSIX Operating Systems. *29th Int. Symp. on Fault Tolerant Computing*. Pages: 30–37. 1999.
- [5] K. Chao and L. Zhang. *Seq. Comparison: Theory and Methods*. Springer, 2009.
- [6] F. Salfner, M. Lenk, and M. Malek. A Survey of Online Failure Prediction Methods. *ACM Computing Surveys*. Vol 42(3). Pages: 1–42. 2010.
- [7] G.S. Lemos, E. Martins. Robustness Testing Oracle using a Sequence Alignment Algorithm. *1st Workshop on Software Test Output Validation*. 2010.

8 Gerenciamento de dados de sensores para Sistemas de informação de Biodiversidade

Autores: Ivo Koga e Claudia M. B. Medeiros

A pesquisa em biodiversidade é um campo multidisciplinar que requer a cooperação de vários tipos de pesquisadores. Os biólogos realizam diferentes tipos de atividades, incluindo coletas em campo, análises de dados sobre os espécimes coletados, seus habitats e correlações com outros seres vivos, construindo modelos capazes de descrever essas interações [1, 2].

Os dados disponíveis vêm sendo coletados em vários lugares do mundo por muitos grupos de pesquisadores, sendo publicados em formatos distintos e especificados em inúmeros padrões. Este cenário é caracterizado por sua heterogeneidade intrínseca – não apenas de dados e modelos conceituais utilizados, como também de necessidades e perfis dos especialistas que coletam e analisam os dados.

Apesar da existência de uma enorme quantidade de dados de biodiversidade publicados na Web, um dos problemas atuais é integrá-los a dados de sensores coletados no meio ambiente (por exemplo, umidade, temperatura).

Até agora há pouquíssimos trabalhos integrando dados de sensores a dados de biodiversidade. Portanto, a investigação de métodos de gerenciamento e integração destes dados, com possibilidade de visualização e manipulação destes dados na Web é importante.

O objetivo do trabalho é especificar e desenvolver um *framework* computacional que permita aos pesquisadores em biodiversidade combinar consultas a dados de coletas com dados obtidos de redes de sensores.

O trabalho pressupõe que os dados de coletas estão distribuídos em repositórios na Web. Cada repositório é mantido por um grupo de cientistas e seu conteúdo é disponibilizado por Serviços Web.

Além das consultas aos repositórios das coletas, serão desenvolvidos mecanismos eficientes para acesso a dados de sensores. Os dados dos sensores fornecerão meios para que pesquisadores de biodiversidade estudem as sutis variações temporais e espaciais dos parâmetros ambientais.

Este trabalho se integrará na arquitetura do BIOCORE [4], projeto do edital CT-INFO 2007 financiado pelo CNPq, que visa a especificação e desenvolvimento de ferramentas computacionais para aos cientistas em biodiversidade gerenciar e compartilhar seus dados, auxiliando-os na construção de modelos complexos e na análise e modelagem de ecossistemas, incluindo a descoberta de novos relacionamentos e interações entre espécies. Esta integração considerará os dados de sensores na arquitetura do BIOCORE permitindo o armazenamento e integração com os outros dados coletados nesse projeto.

Adicionalmente, as pesquisas desenvolvidas durante o doutorado utilizarão como base inicial o trabalho desenvolvido no doutorado de Pastorello [6]. O trabalho irá igualmente considerar aspectos de metadados de sensores [7], para permitir que os dados coletados sejam acessados e combinados aos dados de coletas de biodiversidade.

Por fim será especificado e desenvolvido um *framework* capaz de combinar as consultas dos dados das coletas em biodiversidade ao uso dos dados de redes de sensores. Este irá ajudar os pesquisadores em biodiversidade, já que possibilitará uma avaliação mais

detalhada das características ambientais dos habitats onde foram realizadas as coletas. Isto possibilitará a realização de avaliações mais precisas e correlações entre os diferentes dados dos sensores e das coletas.

Os seguintes resultados deverão ser obtidos do projeto:

- avaliação dos serviços que disponibilizam dados das coletas;
- criação de uma infra-estrutura para recuperação e disponibilização de dados de redes de sensores;
- especificação, com os biólogos do Museu de História Natural da UNICAMP, de um conjunto básico de dados a serem utilizados;
- especificação de um *framework* genérico para dar apoio a consultas, na Web, a dados heterogêneos de biodiversidade e de redes de sensores;
- implementação, testes e validação do *framework*, a partir de um protótipo usando como base o Museu Virtual de História Natural da UNICAMP.

Referências

- [1] T. M. Lewinsohn and G. J. Shepherd. Taxonomic knowledge bases as tools for biodiversity research in the Third World. Proc. VI International Congress of Ecology, Manchester, England, 1994
- [2] T. M. Lewinsohn and P. I. K. L. Prado and V. N. Solferini et al. “How feasible is a full-fledged model system for communities? Charting associations of Brazilian Asteraceae and their flowerhead feeders”. The Role of Model Systems in Ecological Research (10th Bodega Conference), pages 31–35 California, USA, 1997.
- [3] R. S. Torres and C. B. Medeiros and M. A. Gonçalves and E. A. Fox A Digital Library Framework for Biodiversity Information Systems. International Journal on Digital Libraries, 6(1):3–17.
- [4] C. B. Medeiros and R.S. Torres and R. Azevedo and A.X. Falcao and T. Lewinsohn and P.I. Prado and E. Martins and C. Rubira and R. M. Cesar and M. A. C. Amaral and A. Santanchè. BioCORE – ferramentas, modelos e técnicas para apoio a pesquisa em biodiversidade. <http://www.lis.ic.unicamp.br/projects/bio-core-tools-models-and-techniques-to-support-research-in-biodiversity/> (Dec 07).
- [5] Gupchup, J. and Musaloiu-E, R. and Terzis, A. and Szalay, A. and Szlavec, K. and Chang, M. Deploying Advanced Wireless Sensor Networks for Ecological Monitoring. Microsoft eScience Workshop, December, 2008, pages 88–90.
- [6] G. Z. Pastorello Jr Gerenciamento do ciclo de vida de dados de sensores: da produção ao consumo. IC – UNICAMP, December, 2008, Orientador C. M. B. Medeiros.

- [7] N. Dawes and K. A. Kumar and S. Michel and K. Aberer and M. Lehning Sensor Metadata Management and its Application in Collaborative Environmental Research Proc. 4th IEEE eScience Conference, 2008, pages 143–150.

9 Managing Data Provenance for Biodiversity Information

Autores: Joana E. Gonzales Malaverri e Claudia M. Bauzer Medeiros

wadays, scientists from several domains (e.g., bioinformatic, environmental sciences, agriculture, social sciences) are using different kinds of techniques and methodologies to capture, produce, process and analyzes a huge volume of data. This has given rise demands that requires new computational tools to systematically tackle with distributed and heterogeneous data source, data storage, transfer mechanisms, among others, thus providing better cooperative and supporting tools. Common problems among scientific domains are related to how to document and preserve the experiments ensuring its reproducibility; how to organize and integrate it; how to share the findings and how to assess the quality of the data results.

Biodiversity research is a good example that needs to face with these problems. Data collected are manipulated by a wide range of users, with distinct research interests, using its own vocabularies, work methodologies, models, and sampling needs. Hence, data analysis requires the combination of expertise in many scientific fields such as ecology, genetics, and so on. One possible approach to tackle these problems is to elicit knowledge by means a good documentation process. This work proposes a framework to support management of data provenance for biodiversity studies. We study provenance as a means to help end-users to assess the trustworthiness of a scientific result and the processes that generated it.

Figure 1 outlines our architecture, where each box denotes different data access and manipulation levels. The two boxes outlined (encapsulating the Provenance Extraction and Data Quality Assessment processes) are the main focus of this research. The Provenance Extraction process is in charge of extracting the provenance information of the data input, the processes performed and the results obtained. The major issue to be dealt with in this point is related to what may be considered as provenance information and how to capture it. The Data Quality may be invoked to assess result quality, based on provenance information. Part of the other elements are already implemented, as discussed in [1], [2], [4], [3].

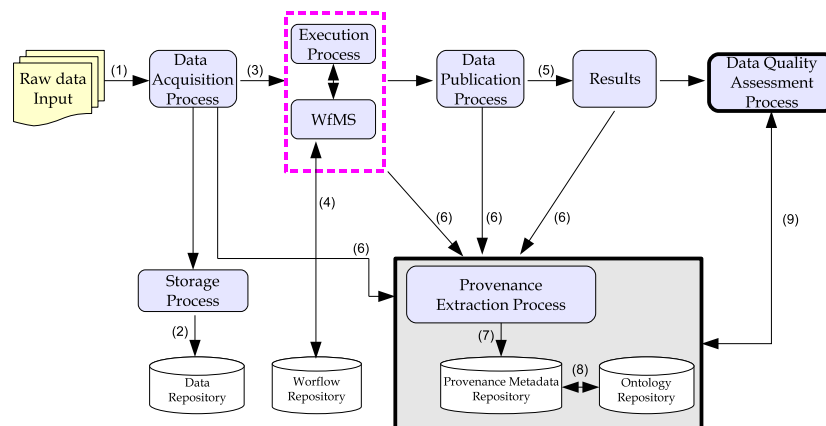


Figure 1: Supporting provenance extraction and management

The challenges associated with this research covers theoretical and implementation issues involving studies which use assessment of data quality, management of data provenance, metadata standards and semantic annotations. This research aims to provide a framework that will support handling provenance issues in biodiversity research. The provenance that we considerate is related to scientific activities - inside or outside of the laboratory - using several methodologies and information sources, involved in this domain, for example habitat characterization, species identification addressed to biological collections, among others.

The provenance information will be used to help the specialists in the assessment of the results generated in a scientific study. The main expected contributions are: (i) the specification of a framework to manage data provenance in the context of web-based biodiversity systems; (ii) the documentation of scientific work through the generation and manipulation of their provenance; (iii) the assessment of the results obtained using its provenance information; and (iv) the definition of requirements to create a model to represent and store provenance, along with the means to access and trace this information. Our framework will consider the integration of heterogeneous data, quality issues and the queries that will be performed on the provenance information.

This work has the following publication: *J. G. Malaverri and C. B. Medeiros. Handling Provenance in Biodiversity. Workshop on Challenges in eScience (CIS-2010).*

Referências

- [1] C. B. Medeiros, J. Pérez-Alcazar, L. Digiampietri, G. Z. Pastorello Jr., A. Santanchè, R. T. da Silva, E. Madeira, and E. Bacarin. WOODSS and the Web: Annotating and Reusing Scientific Workflows. *SIGMOD Record*, 34(3):18–23, 2005.
- [2] C. N. Mac-ri. *Semantic Annotation of Geospatial Data*. PhD thesis, Instituto de Computação - Unicamp, 2009.
- [3] J. E. G. Malaverri, B. Vilar, and C. B. Medeiros. A Tool Based on Web Services to Query Biodiversity Information. In *5th International Conference on Web Information Systems and Technologies (Webist 2009)*, 2009.
- [4] J. G. Malaverri. Um Serviço de Gerenciamento de Coletas para Sistemas de Informação de Biodiversidade. Master's thesis, Instituto de Computação - Unicamp, 2009.

10 Paralelização de Simuladores Descritos em *SystemC*

Autores: Liana Dessandre Duenha e Rodolfo Jardim de Azevedo

SystemC é uma Linguagem de Modelagem e Simulação de Sistemas (SLDL) baseada em C++ utilizada por projetistas de SoCs (*System on a Chip*) e processadores embarcados [2]. Uma limitação desta SLDL é que seu *kernel* de simulação sequencializa a execução das *threads* e não aproveita todos os processadores disponíveis. Por exemplo, um sistema *multicore* modelado em *SystemC* executando em uma plataforma com mais de um processador, seria simulado usando apenas um dos núcleos de processamento. Investigar técnicas para dar suporte à paralelização de simuladores descritos em *SystemC* em busca da aceleração da simulação é um desafio importante e foco de diversas pesquisas em desenvolvimento no meio acadêmico [1, 5, 4].

Este projeto tem como objetivo geral o desenvolvimento de técnicas para habilitar a simulação paralela baseada em eventos discretos (PDES) de sistemas descritos em *SystemC*, visando melhoria de desempenho e transparência para o projetista do modelo. Segue uma breve explicação sobre algumas técnicas que pretendemos utilizar para atingir tal objetivo.

A primeira abordagem consiste na definição de quais *threads* do simulador podem ou não executar em paralelo e para isso precisamos entender como funciona o *kernel* de simulação. O núcleo *SystemC* é um escalonador que sequencializa a execução de *threads* mapeadas como objetos da classe `sc_thread`. Utilizaremos objetos da classe `sc_dthread` definida por Favari(2010) [3] para mapear as *threads* da simulação. Ao encontrar uma chamada de `wait()`, uma `sc_dthread` libera o processador para que outra *thread* execute e continua apta a ser escalonada posteriormente, permitindo execução concorrente.

Para evitar erros decorrentes da execução concorrente, é necessário verificar se os sinais de entrada de uma *thread* são válidos antes de iniciar sua execução. Se isto ocorre, a *thread* pode ser escalonada para execução concorrente. Senão, um laço de espera é realizado antes da sua execução. A Figura 1 ilustra uma simulação utilizando oito *threads* e os sinais de entrada e saída de cada uma delas.

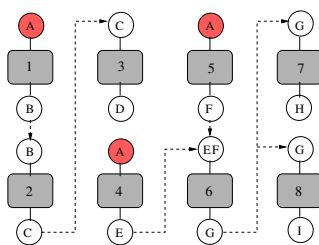


Figura 1

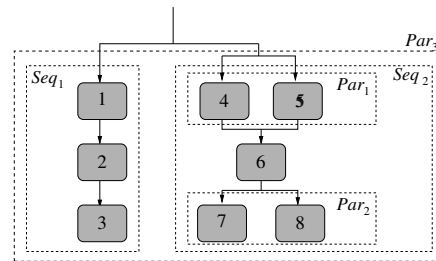


Figura 2

Estudando as dependências entre as *threads*, podemos estabelecer grupos cujas *threads* participantes podem ser executadas em paralelo (Par_1 , Par_2 e Par_3) e grupos que obrigatoriamente devem ter execução sequencial (Seq_1 e Seq_2). A Figura 2 ilustra essa possibilidade de agrupamento e sequência de execução.

Outra abordagem a ser aplicada é a identificação e junção de processos pequenos na

simulação, visando minimizar o tempo gasto com trocas de contexto. Na Figura 3(a) ilustramos a execução de três processos pequenos em um mesmo processador e as várias trocas de contexto realizadas. Na Figura 3(b) ilustramos a criação de um processo P_{12} a partir da junção de P_1 e P_2 em um único processo e uma possível ordem de execução deste processo juntamente com o processo P_3 , diminuindo a quantidade de trocas de contexto.

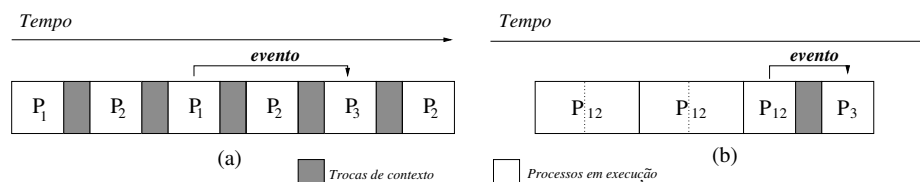


Figura 3

Este trabalho está em fase inicial, portanto não há resultados a serem apresentados no momento. Após o desenvolvimento destas e de outras técnicas, pretendemos aplicá-las em plataformas já desenvolvidas em outros trabalhos de pesquisa da área e nas demais plataformas que desenvolvermos. A avaliação de desempenho será realizada mediante comparação de tempo de processamento e taxa de utilização dos *cores* dos processadores durante simulações com e sem o uso de tais técnicas de paralelização.

Referências

- [1] P. Ezudheen and P. Chandran and J. Chandra and B.P. Simon and D. Ravi. Parallelizing SystemC Kernel for Fast Hardware Simulation on SMP Machines. In *23rd Workshop on Principles of Advanced and Distributed Simulation*, 2009.
- [2] D.C. Black and J. Donovan. SystemC: From the Ground Up Kluwer Academic Publishers, 2004.
- [3] R.R.C. Faveri. Viabilizando a Simulação Multi-Threaded para Modelos Escritos em *SystemC*. Dissertação de Mestrado - IC-Unicamp, 2010.
- [4] B. Chopardl and P. Combes and J. Zory. A Conservative Approach to SystemC Parallelization. In *International Conference on Computational Science*, 2006.
- [5] A. Mello and I. Maia and A. Greiner and F. Pecheux. Parallel Simulation of SystemC TLM 2.0 Compliant MPSoC on SMP Workstations. In *Design, Automation and Test in Europe Conference and Exhibition*, 2010.

11 Arquitetura para Suporte a Aplicações Ubíquas que Viabilizam a Criação de um Ambiente de Aprendizado Ativo em Sala de Aula

Autores: Ricardo Edgard Caceffo e Heloisa Vieira da Rocha

O modelo denominado aprendizado ativo, fundamentado no construtivismo, permite que os alunos tenham um papel ativo em seu processo de aprendizado [7] [1]. Este modelo estimula a construção de um ambiente onde existe uma efetiva interação entre alunos e professor, apoiado pela aplicação de atividades que motivam esse tipo de comportamento.

Atualmente a tecnologia viabiliza o aprendizado ativo através dos sistemas de resposta em sala de aula [2]. Esses sistemas, usualmente fundamentados em dispositivos com suporte à caneta eletrônica (e.g. Tablets PCs), permitem ao professor propor atividades aos alunos, que as respondem em seus dispositivos e as submetem eletronicamente de volta ao professor. Este, por sua vez, é capaz de agregar, analisar e exibir essas submissões aos alunos, o que facilita o feedback, a discussão e a colaboração em sala de aula.

Entretanto, estudos envolvendo o uso dos sistemas de resposta em sala de aula mostram que a aplicação deste tipo de sistema pode gerar perda de foco, atenção e motivação dos alunos [3] [1] [4] [5]. Ainda, o uso específico de um único dispositivo limita a aplicação da tecnologia por não considerar a presença crescente dos mais variados dispositivos no ambiente de ensino.

Entrementes temos o surgimento da computação ubíqua [8], que seria a 3ª onda de evolução da computação, que prevê a existência e interação do usuário com uma série de dispositivos no ambiente. Uma de suas principais características é o conceito da "tecnologia calma" [6], onde os computadores devem ficar em segundo plano, permitindo assim aos usuários o foco nas atividades a serem executadas, minimizando eventuais distrações causadas pelo uso da tecnologia.

Desta forma, é proposta a criação de uma arquitetura para suporte a aplicações ubíquas de sistemas de resposta em sala de aula com o intuito de viabilizar o ambiente de aprendizado ativo. Foram definidos módulos de interação dos usuários, como proposta de atividades, recebimento de submissões e análise e feedback. No momento estão sendo definidas as características técnicas da arquitetura, como modelo de banco de dados, categorização e acesso à informação das atividades e implementação das características associadas à ubiquidade do sistema.

A avaliação será feita através do desenvolvimento de protótipos funcionais a serem aplicados num ambiente real de ensino. Os protótipos estão sendo desenvolvidos para terem suporte a notebooks, Tablets PCs e smartphones, sensíveis ou não ao toque.

Educacionalmente pretende-se verificar se a arquitetura realmente suporta o ambiente de aprendizado ativo, viabilizando a colaboração entre alunos e professor, o feedback e a realização de exercícios e tarefas tanto através da interação pública como privada com os dispositivos. Ainda, pretende-se verificar de qual forma o uso específico de características associadas à computação ubíqua, como a análise do contexto, a pró-atividade e a adaptabilidade do sistema afetam o ambiente de ensino e aprendizado em sala de aula.

Já na esfera computacional espera-se que os protótipos desenvolvidos obedeçam aos

princípios de usabilidade e sejam factíveis de serem aplicados em um ambiente de ensino, composto pelos mais diversos dispositivos móveis, com diferentes protocolos, formas de comunicação entre si e interfaces de interação distintas com os usuários.

Referências

- [1] Anderson. R. Supporting Active Learning and Example Based Instruction with Classroom Technology . In *ACM SIGCSE Bulletin Volume 39, Issue 1, Session: Teaching with tablets and inking technologies table of contents*. 2007. pp69-73.
- [2] Vahey. P et.al. Using Handheld Technology to Move between Private and Public Interactions in the Classroom. In *Ubiquitous Computing in Education: Invisible Technology, Visible Impact*. . Research Center for Educational Technology, Kent State University, New Jersey, London. 2007 pp187-211
- [3] Wolfman. S. Understanding and Promoting Interaction in the Classroom through Computer-Mediated Communication in the Classroom Presenter System. Ph.D dissertation, University of Washington. 2004
- [4] Caceffo. Ricardo. Ferramenta de apoio para o aprendizado ativo usando dispositivos com caneta eletrônica. Dissertação de Mestrado. Instituto de Computação, UNICAMP Março de 2009
- [5] Pedro. A. Desenvolvimento e análise de impacto de uma aplicação colaborativa voltada para o aprendizado utilizando interação pen-based. Dissertação de Mestrado. Instituto de Computação - UNICAMP, 2010
- [6] Mostefaoui. S. Advances in Ubiquitous Computing: Future Paradigms and Directions. IGI Global. IGI Publishing. Hershey, New York, 2008
- [7] Bonwell. C Eison. J Active Learning: creating excitement in the classroom. In *ASHE-ERIC Higher Education Report No. 1*. Washington, DC: George Washington University. 1991
- [8] Weiser. M. The computer for the 21st century. In *Scientific American*,265(3):p94-104, September 1991

12 Geração Automática de Teste Baseado em Modelo

Autores: Thaise Yano e Eliane Martins

A Engenharia de Software recomenda várias técnicas e métodos para lidar com o desenvolvimento de aplicações complexas. Uma delas é baseada no uso de modelos, pois estes permitem abstrair certos aspectos do sistema, tornando mais fácil a sua representação, a comunicação entre os participantes, e também a validação, avaliação e documentação dos sistemas, entre outras. O uso de modelos em testes, ou MBT (do inglês *Model Based Testing*), já vem sendo proposto de longa data, para permitir a automação da geração dos testes. Com o uso de MBT, ao invés de produzir os testes manualmente, como ainda é de praxe, os testes são gerados automaticamente a partir de modelos. Gerar testes manualmente é uma tarefa árdua, demanda esforço e muito tempo, sendo, por isso, altamente propensa a erros, especialmente para aplicações complexas. A abordagem MBT permite que a equipe de testes utilize os mesmos modelos criados pelos desenvolvedores para a criação de testes que cubram variados cenários de uso da aplicação. Uma outra vantagem é que os testes podem ser produzidos uma única vez para serem implementados em diferentes plataformas. Em caso de modificações no comportamento da aplicação, basta atualizar o modelo e gerar novamente os casos de teste, de maneira automática.

O modelo que representa o comportamento do sistema em teste neste trabalho é especificado em Máquina de Estados Finitos Estendida (MEFE). No teste baseado em MEFE, deve-se obter sequências de teste que satisfaçam um critério, como uma transição ou estado que representa uma parte crítica do sistema. A sequência de teste consiste em um caminho no modelo que percorra a transição/estado alvo. Em particular, o critério de teste será a cobertura de transição.

Automatizar a geração de teste contribuiria tanto para melhorar a produtividade quanto para reduzir esforço e custo no processo de desenvolvimento de software. Neste trabalho investiga-se uma abordagem dinâmica para a geração automática de teste baseada em MEFE, descrita em [3]. Na abordagem dinâmica, a geração de teste é reformulada como um problema de otimização. Quando um requisito de teste não é satisfeito, uma função objetivo é associada a ele e métodos de otimização de funções são utilizados para buscar dados que mais se aproximam de satisfazer o requisito. Com base nessas informações, os dados de teste são incrementalmente modificados até que um deles satisfaça o requisito. Neste trabalho é aplicado um algoritmo evolutivo como método de otimização, que utiliza dados de dependências de controle e dados do modelo para guiar a busca. A área de pesquisa que aplica métodos de otimização baseado em busca na geração de teste, tal como os algoritmos evolutivos, é conhecida como teste baseado em busca (em inglês, *Search Based Testing*, SBT) [1, 2]. Algoritmos aleatórios também podem ser usados para explorar o espaço de busca, mas embora fáceis de implementar, não possuem orientação para explorar soluções melhores. Assim, o uso de métodos de otimização é mais promissor, no sentido de que oferecem a possibilidade de balancear entre uma boa investigação do espaço de busca e exploração de melhores soluções. Além disso, o uso de métodos de otimização permite evitar o problema da explosão combinatória, que ocorre com o uso de técnicas exaustivas de busca, no teste baseado em modelo. Na literatura ainda é pouco explorado o uso de métodos

de otimização na geração automatizada de sequências de testes a partir do modelo.

Um dos problemas no teste baseado em MEFE é a geração de sequências de teste ineficazes, que são estruturalmente válidas mas não são executáveis. Isso deve-se ao fato dos métodos de teste apenas considerarem o fluxo de controle da MEFE e não levarem em consideração o fluxo de dados conjuntamente. Devido às condições de guarda das transições, pode haver contradições nos valores dos dados de um caminho e, portanto, a sequência de teste ser ineficaz. Uma condição de guarda pode ser uma expressão lógica ou comparativa que envolvem dados como variáveis e parâmetros, que ao ser satisfeita permite que uma transição seja disparada, mudando o estado do modelo para outro. Para solucionar este problema, o uso de um modelo executável é proposto neste trabalho. Ao invés de gerar testes se baseando na estrutura do modelo, as sequências de teste são dinamicamente obtidas durante a execução do modelo. Para percorrer um caminho no modelo, é necessário satisfazer todas as condições de guarda das transições que formam o caminho. Dessa forma apenas caminhos factíveis são gerados por meio do modelo executável. Além disso, o modelo executável na especificação de requisitos pode ser usado para produzir um protótipo do sistema na fase inicial de desenvolvimento. Esses modelos podem ser utilizados para validar os requisitos do sistema, visto que os usuários podem dar um retorno sobre o comportamento do sistema.

Para avaliar a abordagem proposta foram realizados vários experimentos com modelos da literatura e de aplicações reais. Foi obtida com sucesso uma alta cobertura do requisito de teste nos experimentos.

Referências

- [1] M. Harman, S. A. Mansouri, and Y. Zhang. Search based software engineering: A comprehensive analysis and review of trends techniques and applications. Technical Report TR-09-03, Department of Computer Science, King's College London, April 2009.
- [2] P. McMinn. Search-based software test data generation: a survey. *Software Testing, Verification and Reliability*, 14(2):105–156, 2004.
- [3] T. Yano, E. Martins, and F. L. De Sousa. Most: a multi-objective search-based testing from fsm. In *Proc. SBST'11: 4th Int. Workshop on Search-Based Software Testing*, 2011.

13 Segmentação interativa de objetos em imagens e vídeos baseada em grafos e modelos nebulosos de conhecimento de conteúdo

Autores: Thiago Vallin Spina e Alexandre Xavier Falcão

Um dos desafios de expressar o conteúdo de imagens e vídeos se apresenta na forma de segmentar precisamente seus objetos semânticos (e.g., pessoas, objetos, animais). Soluções automáticas para este problema são frequentemente inviáveis devido à grande heterogeneidade de cenas naturais. Logo, a segmentação é mais factível combinando o conhecimento de alto nível do usuário sobre o objeto de interesse com a capacidade de delineamento preciso da máquina. Portanto, passa a ser necessário maximizar a eficácia da interação do usuário com o sistema, através da minimização da necessidade de sua intervenção na tarefa de segmentação provendo ao mesmo tempo controle total sobre o processo [4].

Neste sentido, métodos para segmentação interativa de imagens têm sido propostos utilizando tanto restrições de região (marcadores) como de borda (pontos âncora). Tais abordagens possuem pontos fortes e fracos complementares, que podem ser explorados para reduzir o envolvimento do usuário. Em trabalho recente propusemos combinar ambos paradigmas em um novo método denominado *Live Markers* (LM), dentro do *framework* da transformada imagem-floresta [2] (*Image Foresting Transform* — IFT) — uma metodologia para o desenvolvimento de operadores de processamento de imagens baseados em conectividade (grafos). *Live Markers* combina a IFT diferencial via competição de sementes [1] (*Differential IFT with Seed Competition* — DIFT-SC) com o popular *live-wire-on-the-fly* [3] (LWOF), aproveitando suas melhores características para eliminar suas fraquezas. Enquanto a DIFT-SC lida com silhuetas complexas de modo eficaz, apresentado porém vazamentos em partes fracas da borda, o LWOF provê segmentações com bordas mais suaves e é capaz de bloquear o vazamento da DIFT-SC, mas requer maior interação e tempo do usuário. LM transforma cada segmento de borda calculado pelo LWOF entre dois pontos âncora em marcadores internos e externos para a DIFT-SC, de forma a bloquear os vazamentos na segmentação calculada pela segunda. A abordagem híbrida do LM permite a segmentação da imagem em tempo linear na primeira iteração e correções em tempo sublinear nas subsequentes, reduzindo ainda a necessidade de intervenções do usuário.

No caso de vídeo, propomos abordar as questões supracitadas através do estudo e desenvolvimento de métodos baseados em modelos dinâmicos de conhecimento de conteúdo que substituam o usuário em tarefas de reconhecimento do objeto no restante do vídeo, após a segmentação interativa do objeto em um quadro inicial. Mais precisamente, estamos interessados em criar modelos nebulosos (*fuzzy*), tanto 2D como 3D, que explorem o conhecimento adquirido com a segmentação em quadros anteriores para executar a segmentação do quadro atual. Um candidato ideal para esta atividade é o modelo das nuvens (*Object Cloud Model* — OCM) [5], que trata reconhecimento e delineamento de forma sinérgica. Este modelo define, para cada posição no quadro, uma região de incerteza onde a borda do objeto deve estar localizada, enquanto o delineamento é feito utilizando a IFT.

O grande desafio é descobrir maneiras de criar OCMs em vídeo digital, visto que tais modelos necessitam de um conjunto de imagens de treinamento composto por segmentações

do objeto em diferentes poses, e como atualizá-los dinamicamente ao longo do vídeo. Mais ainda, queremos tratar problemas como auto-occlusões e objetos que giram em torno de si próprios (e.g., bailarinas e dançarinos) encontrando a projeção bidimensional de uma nuvem 3D — elaborada a partir de modelos genéricos tridimensionais do objeto — que melhor segmente o objeto em um quadro do vídeo. Por fim, o *framework* de segmentação a ser desenvolvido será validado no contexto de edição de imagens e vídeos (i.e., composição por *alpha matting* [6]).

Referências

- [1] A.X. Falcão and F.P.G. Bergo. Interactive volume segmentation with differential image foresting transforms. *IEEE Trans. on Medical Imaging*, 23(9):1100–1108, 2004.
- [2] A.X. Falcão, J. Stolfi, and R.A. Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(1):19–29, 2004.
- [3] A.X. Falcão, J.K. Udupa, and F.K. Miyazawa. An ultra-fast user-steered image segmentation paradigm: Live-wire-on-the-fly. *IEEE Trans. on Medical Imaging*, 19(1):55–62, Jan 2000.
- [4] A.X. Falcão, J.K. Udupa, S. Samarasekera, S. Sharma, B.E. Hirsch, and R.A. Lotufo. User-steered image segmentation paradigms: Live-wire and live-lane. *Graphical Models and Image Processing*, 60(4):233–260, Jul 1998.
- [5] Paulo A. V. Miranda, Alexandre X. Falcão, and Thiago V. Spina. Riverbed: A Novel User-Steered Image Segmentation Method Based on Optimum Boundary Tracking. Technical Report IC-11-04, Institute of Computing, University of Campinas, January 2011.
- [6] J. Wang and M. Cohen. Image and video matting: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(2):97–175, 2008.

14 Método Ágil aplicado ao Desenvolvimento de Software Confiável baseado em Componentes

Autores: Alan Braz, Cecília M. F. Rubira e Marco Vieira

A popularização do Método Ágil (ou Desenvolvimento Ágil de Software - DAS)[4] através de metodologias como XP e Scrum fez com que fossem aplicadas no desenvolvimento de sistemas computacionais robustos. Este fato evidencia a necessidade de processos de desenvolvimento e gerenciamento de software que sejam robustos e que possuam uma quantidade adequada de modelagem e documentação, em especial no que concerne o projeto arquitetural, a fim de garantir maior qualidade no seu resultado final. A robustez pode ser alcançada adicionando elementos de tratamento de exceções às fases iniciais do processo de desenvolvimento e na reutilização de componentes.

O objetivo deste trabalho é desenvolver a combinação do método que auxilia a construção de sistemas de software com requisitos de confiabilidade ligados à tolerância a falhas, chamado de MDCE+[1], ao processo de DAS Scrum[5]. Isso será feito através da utilização das tarefas de identificação de exceções e a especificação de seus tratadores nas fases do Scrum. O método resultante será chamado de Scrum+CE (Scrum com Comportamento Excepcional) e possibilitará o desenvolvimento de software confiável utilizando um método ágil.

O Scrum+CE foi definido com as alterações que serão necessárias nas fases de Pré-jogo e Jogo no que diz respeito a levantar Histórias Excepcionais (HE) e refinar a Arquitetura com elementos excepcionais explícitos. Além disso, técnicas ágeis que melhoram a confiabilidade como *Test-Driven Development* (TDD) e Integração Contínua (IC) são fortemente recomendadas.

A Figura 2 mostra as fases do Scrum com as atividades do MDCE+ (em cinza) que devem ser desenvolvidas nas respectivas fases. As maiores modificações serão na fase de Pré-Jogo, na qual o comportamento excepcional será documentado na forma de Histórias Excepcionais e de Testes de Aceitação nas Histórias de Usuário (HU)[1] que validem especificamente os fluxos excepcionais. Além disso, a Arquitetura em alto-nível passará a expor os componentes excepcionais conforme o modelo de componente tolerante a falhas ideal que é categorizado em dois tipos: normal, que produz respostas corretas; e excepcional (ou anormal), que é executado quando um erro é detectado.

Dentro da fase Jogo, teremos como entradas as HU e HE igualmente priorizadas no *Backlog do Produto*. Elas deverão entrar no *Backlog do Sprint* de acordo com suas prioridades e dependências.

A avaliação dos resultados será feita em projetos selecionados para um estudo de caso qualitativo revisando os requisitos e o documento de arquitetura, adaptando-os as práticas e técnicas do Scrum+CE.

Será feita uma análise dos benefícios esperados e alcançados usando-se questionários aplicados aos desenvolvedores dos projetos e comparando as estimativas em Pontos de História (*Story Points*)[3].

Após a adaptação e validação espera-se um aumento no número de HU, e por consequência de Pontos de História, acarretando no aumento no tempo de desenvolvimento e

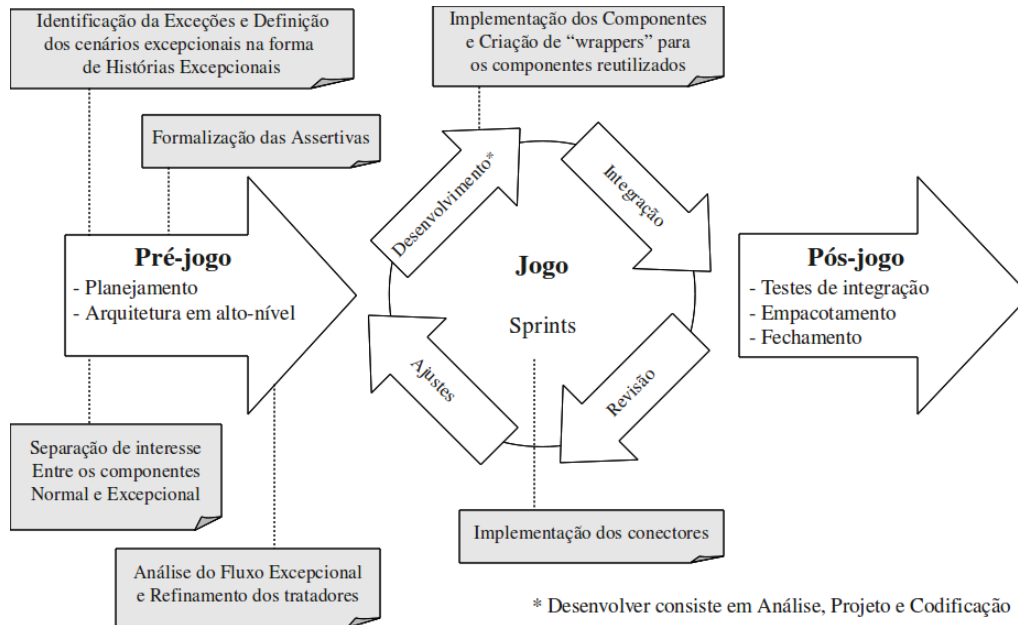


Figura 2: Interferência do Método MDCE+ nas fases do Scrum

custo. Em contrapartida espera-se a diminuição dos defeitos na fase de Pós-Jogo.

Referências

- [1] Brito, P. H. S. (2005). *Um Método para Modelagem de Exceções em Desenvolvimento Baseado em Componentes*. Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Campinas, Instituto de Computação. Orientador: Cecília Mary Fischer Rubira.
- [2] Cohn, M. (2004) *User Stories Applied: For Agile Software Development*, Addison-Wesley.
- [3] Cohn, M. (2005) *Agile Estimating and Planning*, Prentice Hall.
- [4] Manifesto Ágil (2001). *Manifesto para o desenvolvimento ágil de software*. Disponível em: <http://manifestoagil.com.br/>. Acessado em: 17 jan 2011.
- [5] Schwaber, K. (1995). *SCRUM Development Process*, OOPSLA'95 Workshop on Business Object Design and Implementation. Austin, USA. Disponível em: <http://www.jeffsutherland.org/oopsla/schwapub.pdf>. Acessado em: 22 mar 2010.

15 Análise Comportamental de Código Malicioso Através da Monitoração de Chamadas de Sistema e Tráfego de Rede

Autores: Dario Simões Fernandes Filho, Paulo Lício de Geus

Nos dias de hoje a internet vem tornando mais pratica o dia a dia das pessoas devido ao grande número de serviços disponíveis. Grande parte deles envolvem operações financeiras como por exemplo compra e venda de produtos e *internet banking*. Essa praticidade no entanto atrai um tipo indesejado de *software*, os *malware*. Normalmente este tipo de programa realiza ações na máquina do usuário sem o seu consentimento podendo roubar e destruir dados importantes [1].

De acordo com pesquisas recentes, o número de *malware* observados no período entre janeiro a setembro de 2010 foi de aproximadamente 14 milhões, um valor superior em um milhão ao mesmo período do ano passado. [2]. Isto torna indispensável o uso de medidas de combate a este tipo de problema. Um delas é o uso de antivírus, *software* que ficam monitorando a máquina constantemente em busca de arquivos infectados ou com comportamento suspeito. Para identificar estes arquivos é necessário primeiro ter um exemplo de comportamento malicioso, que pode ser obtido através da análise dinâmica de artefatos maliciosos. Com ela é possível verificar quais ações foram realizadas pelo programa durante a sua execução. O resultado gerado por essa análise possibilita que seja obtido o comportamento malicioso do *malware*.

O trabalho proposto visa criar um sistema de análise de artefatos maliciosos (BEHEMOT) que consiga obter informações sobre a execução do artefato e o tráfego de rede gerado durante a análise. Esse sistema deve ser capaz de analisar arquivos do tipo PE 32 *Executable* executáveis de sistemas *Windows*. Como grande parte dos computadores usam *Windows* [3] a maioria dos *malware* são desenvolvidos para este sistema operacional. Além disso, ele não deve ser dependente de arquitetura como alguns sistemas que só executam em ambiente emulado como por exemplo o Anubis [4], um sistema de análise dinâmica muito conhecido e utilizado que utiliza uma versão modificada do Qemu [5] para realizar a análise do *malware*. Essa característica é importante pois o sistema deve ser utilizara arquitetura emulada e real, tornando possível a análise de artefatos que não podem ser executados em ambiente emulado. Outra característica desejada é com relação aos resultados das análises. Eles devem apresentar somente as ações maliciosas efetuadas pelo binário, excluindo o máximo possível qualquer ruído proveniente de ações do sistema operacional.

Os resultados preliminares do BEHEMOT mostram que ele consegue analisar mais *malware* que outras ferramentas, devido sua arquitetura híbrida que utiliza máquinas reais e emuladas na análise. Outro ponto relevante foram as análises que se mostraram mais objetivas, com poucas ações provenientes do sistema operacional apresentando em sua maioria somente o comportamento malicioso.

Referências

- [1] P. Vinod, V.Laxmi M.S.Gaur. Survey on malware detection methods. <http://www.security.iitk.ac.in/contents/events/workshops/iitkhack09/papers/vinod.pdf>.

- [2] McAfee. McAfee Labs. McAfee threats report: Third quarter 2010. Technical report. http://www.mcafee.com/us/local_content/reports/q32010_threats_report_en.pdf, 2010.
- [3] W3Schools. Estatística sobre uso de sistemas operacionais. http://www.w3schools.com/browsers/browsers_os.asp., 2011
- [4] Ulrich Bayer, Christopher Kruegel, Engin Kirda. TTanalyze: A Tool for Analyzing Malware. In *Proceedings 15th Ann. Conf. European Inst. for Computer Antivirus Research (EICAR)*, pages 180–192, 2006.
- [5] Fabrice Bellard. QEMU, a fast and portable dynamic translator. In *In Proceedings of the Annual Conference on USENIX Annual Technical Conference*, pages 41–41, 2005.

16 Integrating and Enriching Folksonomies and Ontologies

Authors: Hugo Alves, and André Santanchè

The popularization of web-based systems offering services for content storage, indexation and sharing fostered a rapid growth of content available on-line. There are more than 5 billion images hosted on Flickr² and more than 180 million URL addresses on Delicious³. These systems increasingly rely on tag-based metadata to organize and index all the amount of data.

The tags are provided by users usually connected in social networks, who are free to use any word as tag; there is no central control. Tagging in a social environment has been called folksonomies, combining the words “folk” and “taxonomy” [8].

Any operation involving indexation, classification or discovery will require comparison among the involved tags. In this topic, there are approaches ranging from a pure lexical or statistical comparison of words to a richer semantic analysis of relations, by associating tags to formal ontologies. In many contexts, this semantic directed approach would enable machines to better classify, rank, disambiguate, and discover tags, enriching the systems and the user experience. However, in front of the easiness offered by tag producing, there is a discrepant extra effort to relate tags to ontologies. Recent studies show that folksonomies store potential semantics, which can emerge by the comparison of relations held among tags [2]. It fostered proposals to automatically derive ontologies from folksonomies [1, 3, 6, 7]. Even showing promising improvements, we believe that this semantics derived from folksonomies will achieve better results by relating it to preexisting ontologies.

Therefore, our approach, presented in this work, takes advantage from the best of both worlds – ontologies and folksonomies. In one direction, previously built ontologies are explored to enhance operations over folksonomies. On the other direction, we extract semantics from the relations found in folksonomies to enrich ontologies with statistical data, which will improve ontology-based operations, e.g., analysis of semantic similarity.

We validated our approach in a practical experiment involving more the 50000 tags collected from Flickr and Delicious, which are related to the WordNet ontology [4]. Our solution was compared to other approaches in order to show its improvements.

Referências

- [1] Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic grounding of tag relatedness in social bookmarking systems. In *The Semantic Web – ISWC 2008, Proc.Intl. Semantic Web Conference 2008*, volume 5318 of *LNAI*, pages 615–631, Heidelberg, 2008. Springer.
- [2] Thomas Gruber. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, 3(2):1–11, 2007.

²<http://blog.flickr.net/en/2010/09/19/5000000000/>

³<http://blog.delicious.com/blog/2008/11/delicious-is-5.html>

- [3] Ioannis Konstas, Joemon M Jose, and Ivàn Cantador. Categorising social tags to improve folksonomy-based recommendations. *World Wide Web Internet And Web Information Systems*, 9(1):1–15, 2010.
- [4] G. A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.
- [5] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In C. Raymond Perrault, editor, *Proceedings of the 14th IJCAI*, pages 448–453, Montréal (Canada), 1995.
- [6] L. Specia and E. Motta. Integrating folksonomies with the semantic web. In *Proceedings of the European Semantic Web Conference (ESWC2007)*, volume 4519 of *LNCS*, pages 624–639, Berlin Heidelberg, Germany, July 2007. Springer-Verlag.
- [7] Céline Van Damme, Martin Hepp, and Katharina Siorpaes. Folksontology: An integrated approach for turning folksonomies into ontologies. In *Proceedings of the ESWC Workshop “Bridging the Gap between Semantic Web and Web 2.0”*. Springer, 2007.
- [8] Thomas Vander Wal. Folksonomy. <http://vanderwal.net/folksonomy.html>, 2007. Retrieved on April, 2011.

17 Explicitação de Esquema Orientada a Contexto para Promover Interoperabilidade Semântica

Autores: Ivelize Rocha Bernardo, André Santanchè

Grande parte da informação disponível no mundo está em planilhas eletrônicas [1], por isso há uma forte preocupação, tanto no setor empresarial quanto no acadêmico, no sentido de tornar estes dados mais compartilháveis e flexíveis [2], de forma que outros aplicativos possam também interpretar os dados e os resultados gerados.

Deste modo, a interoperabilidade é um foco central das pesquisas. Do ponto de vista sintático, diversos avanços já foram realizados, seja por padrões mais simples como o CSV (Comma Separated Values), quanto por aqueles mais complexos, a exemplo dos baseados em XML. Por outro lado, a interoperabilidade semântica ainda é objeto de debate, as abordagens para alcançá-la variam desde processos de mapeamento manual para padrões abertos da web semântica, até propostas para reconhecimento automático de estruturas pela associação dos elementos da planilha a conceitos disponíveis em bases de conhecimento da web – e.g., DBpedia (<http://dbpedia.org>).

Verificamos nas abordagens analisadas que o processo de explicitação do esquema presente em uma planilha é desvinculado de um reconhecimento prévio do seu contexto. Acreditamos que uma caracterização deste contexto pode guiar tal reconhecimento, gerando resultados mais consistentes e com mais riqueza semântica.

Este trabalho visa contribuir nesta direção. Ele parte de uma pesquisa prévia denominada Semântica In Loco [3], que envolve, entre outras coisas, o reconhecimento de esquemas implícitos em documentos e a sua transformação em formatos segundo padrões abertos de interoperabilidade. Nossa proposta é estender a Semântica In Loco, de modo que ela utilize informações de contexto como suporte à interpretação dos esquemas implícitos contidos nas planilhas eletrônicas.

Deste modo, partiremos da noção de contexto proposta por [4]: “Contexto é qualquer informação que pode ser usada para caracterizar uma situação ou entidade.” (tradução livre). Especificamente, será usada uma estratégia de reconhecimento de contexto a partir do modelo que dá origem à planilha. O processo proposto está baseado em uma sequência de ações, que estão organizadas em três etapas:

- criação, documentação e armazenamento de modelos;
- explicitação de esquemas baseada em contexto;
- enriquecimento semântico a partir do contexto explicitado.

Com isto, a contribuição central desta pesquisa envolve aprimorar o processo de interpretação de esquemas implícitos justamente pela utilização desses dados de contexto. Acreditamos que a identificação do contexto do usuário orientará um processo mais eficiente de interpretação dos esquemas implícitos contidos nestas planilhas, bem como uma melhor conversão para padrões da web semântica.

Referências

- [1] Syed, Z., Finin, T., Mulwad, V., Joshi, A. Exploiting a Web of Semantic Data for Interpreting Tables. In *Proceedings of the Second Web Science Conference*, 2010.
- [2] Zhao, C., Zhao, L., Wang, H. A Spreadsheet System based on Data Semantic Object. *IEEE International Conference on Information Management and Engineering (ICIME)*, Chengdu, China, pp. 407 – 411, 2010.
- [3] Santanchè, A. Otimizando a Anotação de Objetos de Aprendizagem através da Semântica In Loco. In *Anais do XVIII Simp. Brasileiro de Informática na Educação*, pp. 526–535, 2007.
- [4] Dey, A. K. Understanding and Using Context. *Personal Ubiquitous Comput* 5(1), 4-7. London, UK: Springer-Verlag, 2001.

18 Complexidade de Construção de Árvores PQR

Autores: João Paulo Pereira Zanetti e João Meidanis

Apresentamos aqui as árvores PQR, estruturas de dados usadas para resolver o problema dos uns consecutivos, assim como problemas relacionados a ele.

O problema dos uns consecutivos tem relevância em diversas áreas do conhecimento e pode ser definido como segue: dados um conjunto U , de n elementos, e uma coleção $\mathcal{C} = \{C_1, C_2, \dots, C_m\}$ de m subconjuntos de U , chamados de *restrições*, o algoritmo deve decidir se existe uma permutação dos elementos de U em que os elementos de cada $C \in \mathcal{C}$ apareçam consecutivamente. Chamamos tal permutação de *válida*.

Em 1976, o primeiro algoritmo linear para resolver este problema foi desenvolvido por Booth e Lueker [1]. Seu algoritmo constrói incrementalmente uma estrutura de dados chamada de árvore PQ, que representa de forma compacta todas as permutações válidas.

Entretanto, duas inconveniências fazem com que consideremos a árvore PQ a ferramenta definitiva para lidar com o problema dos uns consecutivos. Em primeiro lugar, não existe árvore PQ para entradas que não tenham a propriedade dos uns consecutivos, isto é, quando não há permutação válida. Isto faz com que informações dos estados anteriores da árvore, que poderiam ser úteis para algumas aplicações, sejam perdidas. Outra desvantagem é que o algoritmo de Booth e Lueker é de difícil implementação.

Assim, vários autores propuseram posteriormente outras soluções para o problema, como é o caso das árvores PQR, a alternativa em que concentramos nossos estudos. Esta estrutura é uma generalização das árvores PQ, introduzida por Meidanis e Munuera em 1996 [4]. As árvores PQR têm um tipo adicional de nó, o nó R, que indica que não existe permutação válida para a entrada dada. Assim, as árvores PQR representam, além das permutações válidas, se existirem, todas as restrições dadas e mais as que derivam delas através de algumas operações sobre conjuntos.

Além disso, as árvores PQR tencionam também ser de fácil implementação em relação às árvores PQ. Enquanto o algoritmo de Booth e Lueker exige a verificação de nove padrões de substituição, o algoritmo online para a construção de árvores PQR utiliza seis regras de transformação [6, 7].

Ao estudar o algoritmo online de Telles e Meidanis, encontramos um problema quanto ao uso de estruturas de union-find, que são utilizadas para melhorar a eficiência de algumas operações e permitem a complexidade quase-linear. As estruturas utilizadas em algoritmos eficientes de union-find não permitem que um nó seja removido de um conjunto. Descobrimos, porém, que isto poderia acontecer em alguns casos, o que impossibilitaria o uso de union-find como esperávamos. Uma solução foi encontrada ao criar uma variação de algumas operações para o caso em que um nó teria antes que ser removido.

Também estudamos a possibilidade de um algoritmo offline linear, com base nos algoritmos de Meidanis, Porto e Telles [5] e de McConnell [3]. Este último é linear, mas levantamos uma dúvida quanto a complexidade de um algoritmo de Dahlhaus [2] utilizado por McConnell.

Referências

- [1] K. S. Booth and G. S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *Journal of Computer Systems Science*, 13(3):335–379, 1976.
- [2] E. Dahlhaus. Parallel algorithms for hierarchical clustering and applications to split decomposition and parity graph recognition. *Journal of Algorithms*, 36(2):205–240, 2000.
- [3] R. McConnell. A certifying algorithm for the consecutive-ones property. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, page 777. Society for Industrial and Applied Mathematics, 2004.
- [4] J. Meidanis and E. G. Munuera. A theory for the consecutive ones property. In *Proceedings of WSP'96 - Third South American Workshop on String Processing*, pages 194–202, August 1996.
- [5] J. Meidanis, O. Porto, and G. P. Telles. On the consecutive ones property. *Discrete Applied Mathematics*, 88:325–354, 1998.
- [6] G. P. Telles and J. Meidanis. Building PQR trees in almost linear time. Technical Report 03-26, Institute of Computing, University of Campinas, November 2003.
- [7] J. P. P. Zanetti and J. Meidanis. Construção incremental de árvores PQR. Technical Report IC-10-31, Institute of Computing, University of Campinas, October 2010.

19 Abordagem para Interoperabilidade Semântica em Bases de Dados Multi-Representação

Autores: Luís Theodoro Oliveira Camargo e André Santachè.

A ciência sempre evoluiu a partir da colaboração entre pesquisadores e esse fato fica mais evidente quando tratamos de temas multidisciplinares. Um desses muitos temas está relacionado a pesquisas no contexto de meio ambiente, que pode envolver biólogos, ambientalistas e geólogos.

Pesquisadores desta área, na maioria das vezes, trabalham com dados que representam perspectivas diferentes de um mesmo problema, tendo usualmente modelos de representação diferentes para o mesmo. Por exemplo, um geólogo trabalha com mapas que representam a constituição do solo, enquanto um biólogo pode trabalhar com fotografias de satélite para mapear o tipo de vegetação, esses dois mapas representam a mesma área, mas podem ser representados em modelos e escalas diferentes. Neste contexto, diversas pesquisas se baseiam no princípio da multi-representação. Segundo [3], a multi-representação engloba diferentes perspectivas do observador, em que há variações de escala, resolução e critério de generalização.

Um dos desafios de se trabalhar com multi-representação é a necessidade de se manter a coerência das diversas representações associadas ao mesmo objeto, quando este sofre alterações. Isto se deve ao fato de que usualmente a alteração incidente sobre uma das representações e necessita ser repercutida para as demais. O termo “multiple representation database” (MRDB) se refere a uma estrutura de armazenamento que possui diversas representações de uma mesma entidade ou fenômeno geográfico [2]. Uma das formas de relacionamento entre objetos de um MRDB é derivada da generalização. A generalização no contexto da cartografia consiste na definição de representações simplificadas de elementos, que são compatíveis com escalas ou propósitos específicos de mapas [2]. No processo de simplificação, a generalização oculta detalhes e/ou sintetiza conjuntos de elementos os quais está representando. As possibilidades de compartilhamento e colaboração resultantes do advento da Internet exigem uma ampliação no espectro de ação do que é convencionalmente tratado no domínio da multi-representação. O desafio passa a ser a articulação dos dados não apenas com múltiplas representações, mas também de maneira interoperável. Isto implica em trazer para o contexto de representação abertas da Web, dados disponíveis em bases multi-representadas, que usualmente estão em um formato próprio. Isto permitiria potencializar a ação da multi-representação pela sua associação com Web Semântica, de modo que a semântica associada às abstrações das múltiplas representações possa ser compartilhada. A partir do que foi discutido podemos definir como objetivo dessa pesquisa a criação de uma abordagem para associar descrições semânticas a dados multi-representados. Tal semântica será explorada para manter a coerência entre os dados em suas transformações, além de auxiliar a sua associação a padrões abertos da Web Semântica [1], subsidiando sua interoperabilidade. Para que tal objetivo seja alcançado os seguintes desafios devem ser solucionados:

- **Identificação interoperável de objetos disponíveis em um MRDB:** Um dos passos para compartilhar dados é a definição de um identificador que seja indepen-

dente de implementação, para subsidiar a interoperabilidade. Portanto é desejável que a definição de identificadores seja independente da estrutura de armazenamento. Nossa proposta para resolver essa questão é a utilização de URIs. Essa abordagem vai nos permitir identificar não somente o dado, mas também sua base de origem. Para possibilitar a utilização dessa proposta independente do MRDB, que pode não suportar a utilização de URI como identificador, iremos definir um modelo de criação dinâmica destas URIs a partir dos dados do MRDB.

- **Associação de semântica às estruturas:** Uma vez definidas as URIs para cada objeto, podemos fazer a conexão da URI com os conceitos de uma ontologia. O diferencial dessa proposta é a utilização de duas ontologias, uma de domínio e uma estrutural, que vai permitir descrever os objetos de um MRDB e as transformações entre eles. A ontologia sobre estruturas nos permite criar uma estratégia de verificação de consistência quando um dado sofre uma alteração, permitindo que ela se reflita em todas as representações deste dado.
- **Elaboração da ontologia de transformação:** Para a criação dessa ontologia de transformação iremos utilizar como base trabalhos da área de generalização. A ontologia aqui proposta ainda deverá ser baseada na ontologia estrutural que será definida, pois cada transformação terá um conjunto de estruturas a qual poderá ser aplicada.
- **Aplicações:** As possíveis aplicações da proposta deste trabalho são: (i) garantir a consistência dos dados independente da representação, utilizando-se da semântica dos dados para isto; (ii) permitir inferências semânticas entre os dados e seus relacionamentos; (iii) permitir que MRDBs de domínios diferentes possam ser conectados auxiliando em pesquisas multidisciplinares.

A solução desses desafios nos permitirá desenvolver uma abordagem para descrição semântica de objetos e suas transformações em bancos de dados multi-representação, para subsidiar operações de verificação de consistência e possibilitar inferências sobre estes dados.

Referências

- [1] T. Berners-Lee, J. Hendler and O. Lassila. The Semantic Web. *Scientific American*, May 2001, p. 28-37, 2001.
- [2] L. T. Sarjakoski. Conceptual models of generalisation and multiple representation. *W.A. Mackaness, A. Ruas and L.T. Sarjakoski, Editors, Chapter 2 of generalisation of geographic information: Cartographic modelling and applications, series of international cartographic association*, Elsevier, p. 11 -35. 2007
- [3] S. Zhou, C. B. Jones. A multi-representation spatial data model. In *Proceedings of the 8th International Symposium on Advances in Spatial and Temporal Databases, SSTD 2003, Lecture Notes in Computer Science 2750*, Springer, Berlin, p. 394-411. 2003.

20 Algoritmos Aproximados para Problemas de Empacotamento em Faixa com Restrições de Descarregamento

Autores: Jefferson Luiz Moisés da Silveira⁴ e Eduardo Candido Xavier.

Problemas de Empacotamento aparecem em aplicações práticas e a busca por bons algoritmos para estes é fortemente motivada por razões econômicas. Estes problemas são utilizados em arranjo de caixas em caminhões, problemas de corte de insumos (aço, tecidos e vidros por exemplo) em indústrias, alguns problemas específicos de escalonamento, entre outros. Uma das versões clássicas de problemas de empacotamento é o Problema de Empacotamento em Faixa, onde um conjunto de itens deve ser empacotado em uma única faixa de tal forma a maximizar a área da faixa preenchida pelos itens, ou seja, minimizar a altura total do empacotamento.

Neste trabalho, consideramos o Problema de Empacotamento em Faixa com Restrição de Descarregamento (SPU - *Strip Packing with Unloading Constraints*). Nesta versão do problema, além das informações de largura e altura, cada item possui uma ordem (classe) e o empacotamento final deve ser tal que cada item possa ser “removido” pelo topo da faixa usando apenas a área livre da mesma. Portanto, o empacotamento deve garantir que não seja necessário alterar a organização dos itens no recipiente durante a remoção de um item, ou seja, garantir que exista um caminho livre para o item percorrer entre a sua posição e a saída do recipiente no momento de sua remoção.

Ese problema aparece no contexto de um problema de roteamento chamado 2L-CVRP (*The Vehicle Routing Problem with Two Dimensional Loading Constraints*) [1, 2, 3, 4]. Neste problema, busca-se minimizar o custo do transporte necessário para realizar entregas de produtos a clientes em diferentes localizações. Estes produtos (itens) são enviados em veículos (recipientes) que estão inicialmente situados no fornecedor dos produtos. Para acomodar os itens nos recipientes é utilizado um algoritmo de empacotamento com restrição de descarregamento, para garantir que, durante a entrega de produtos de cada cliente, não haja nenhum produto de outros clientes bloqueando a saída do recipiente.

Na Figura 3 é possível ver dois exemplos de empacotamento para o 2L-CVRP. Na parte (a) é apresentado uma sequência de clientes que serão visitados para entrega de produtos (A-B-C). Em (b) é apresentado um empacotamento inválido, pois não há como remover todos os itens de A (o primeiro cliente) sem movimentar os outros itens do recipiente, considerando que a saída é realizada pelo topo da faixa. Por fim, em (c) temos um empacotamento válido para o SPU.

Em geral, problemas de empacotamento são problemas de otimização que pertencem à classe NP-Difícil. O SPU, por ser uma generalização do Problema de Empacotamento em Faixas clássico, é fortemente NP-Difícil.

Dada a impossibilidade de resolver tais problemas de maneira exata e eficiente, busca-se alternativas que sejam úteis. Na prática vários métodos são utilizados como programação inteira, heurísticas e meta-heurísticas, algoritmos de aproximação (algoritmos aproximados), dentre outros. Dentre os métodos citados, destacamos os algoritmos de aproximação, cujo interesse é, apesar de sacrificar a otimalidade, fazê-lo de forma que ainda possamos dar

⁴Esta pesquisa foi financiada pelo CNPq (Processo 130123/2009-9).

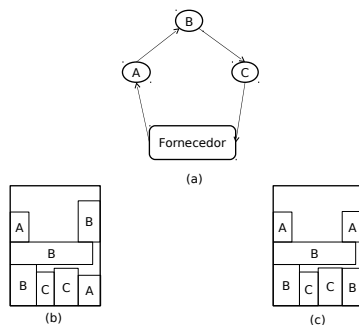


Figura 3: O problema SPU.

boas garantias sobre o valor da solução obtida, procurando ganhar o máximo em termos de eficiência computacional.

Durante o trabalho desenvolvemos 5 heurísticas originais propostas para o SPU. Primeiramente desenvolvemos um novo algoritmo 6.75-aproximado para o problema, baseado no *Bin Packing* bidimensional. Apresentamos também um algoritmo 2-aproximado para o caso específico onde instâncias são tais que o número de classes dos itens é constante (número de clientes na rota). Por fim, apresentamos uma heurística GRASP para o SPU que é fortemente baseada no algoritmo que possui o melhor resultado prático para o Problema de Empacotamento em Faixa clássico [5].

Experimentos computacionais demonstraram que, apesar dos bons resultados obtidos pelos algoritmos aproximados, a heurística GRASP obteve os melhores resultados de maneira consistente. Entretanto, a heurística teve um custo computacional bem mais elevado.

Referências

- [1] M. Iori and J-J. Salazar-Gonzalez e D. Vigo. An Exact Approach for the Vehicle Routing Problem with Two-Dimensional Loading Constraints. *TRANSPORTATION SCIENCE*, 41(2):253–264, 2007.
- [2] M. Gendreau, M. Iori, G. Laporte e S. Martello. A Tabu search heuristic for the vehicle routing problem with two-dimensional loading constraints. *Networks*, 51(1):04–18, 2008.
- [3] E. E. Zachariadis, D. C. Tarantilis e T. C. Kiranoudis. A Guided Tabu Search for the Vehicle Routing Problem with two-dimensional loading constraints. *European Journal of Operational Research*, 195(3):729–743, 2009.
- [4] G. Fuellerer, K. F. Doerner, R. F. Hartl e M. Iori. Ant colony optimization for the two-dimensional loading vehicle routing problem. *Comput. Oper. Res.*, 36(3):655–673, 2009.
- [5] R. Alvarez-Valdes, F. Parreño e J. M. Tamarit. Reactive GRASP for the Strip-Packing Problem. *Computers and Operations Research*, 35:1065–1083, 2008.

21 Shadow-driven Annotation

Authors: Matheus Silva Mota and Claudia Bauzer Medeiros

Despite of the large volume of documents, due to the easy access to richer production tools, formats proliferation is a result of both specific environments and available tools multiplication. In most cases, those tools have not been conceived to produce files with explicit structure and interoperable formats, strongly coupling the content to the file structure and software representation [5, 2]. Furthermore, production tools have offered increasingly support for more than flat text, handling also artifacts like tables, hyperlinks, charts, shapes, embedded multimedia content etc. This likelihood of allowing the construction of richer documents composed of several internal artifacts will help the content comprehension. On the other hand, this capability added to specific formats impacts on the interoperability of those internal artifacts with other systems.

In a scenario with high diversity of inter and non-interoperable formats and a large volume of complex documents, challenges arise when it comes to the systematization of management and storage techniques, retrieval, interpretation and correlation algorithms and new methodologies to present, annotate and mine documents and its internal content, including internal artifacts. In addition, there are problems related to multiple domains – for instance, in the context of scientific research, each area may require different needs of document handling [2].

Document retrieval systems are widely using automatically generated semantic annotations or free manual remarks, made by specialized or general users, on the indexation and retrieval process [3, 4, 1]. Existing tools for manual document annotation can be divided in two categories. The first is the local-based, which produces a local file that contains annotations and/or modify the original document by inserting annotations into it. This local-based approach hamper the process of indexing, exchange and share annotations of a specific document. The second category is the Web-based, which holds documents and annotations on a cloud-based database or library. Since the annotations are produced, those can be shared with a specific group of users or can be used by others applications.

Independently of the storage approach, present annotation tools allows three main annotation strategies. The first strategy concentrate on some specific file format, converting the original document and annotating the converted file. The second strategy requires a file that follow some interoperable standard (e.g., XML). On the other hand, the third strategy deals with documents as images, allowing annotations by floating free-shape layers. Those approaches presents problems when (i) original file preservation is needed; (ii) format diversity handling is needed; and (iii) documents specificity and internal artifacts handling is needed, respectively.

This work presents a novel strategy to annotate documents independently of format, preserving the original file and handling large volume of distributed documents and annotations. This technique also support annotations with multiple granularity levels, allowing annotations of an entire documents or parts thereof (e.g., a specific image, chart, embed multimedia content etc).

The SDA – *Shadow-driven Annotation* – strategy aims to build an interoperable docu-

ment descriptor that summarize the document key-aspects in a XML tree structure. The *shadow* construction is driven by linear reading of the document and its relevant aspects (e.g., pages, paragraphs, embedded multimedia artifacts etc.). Each one of those document key-aspects is mapped into the *shadow* by widely used markup languages (docbook, XSL-FO etc.), metadata sets (e.g., DublinCore, XMP etc.) and Unique Resource Identifiers – URIs. Once the interoperable document descriptor is produced, we are able to use some already established patterns (such as Xpointer, Xlink and RDF) to address and annotate a entire document or fragments of it.

The main advantages of the SDA approach are: (i) the *shadow* generation process is regardless of file type; (ii) the *shadow* summarizes, describes and gives URIs to embedded artifacts and key-aspects of a document, allowing annotation of a entire document or a specific fragment of it; (iii) *shadows* may have different granularity levels, allowing the construction of specialized descriptors for different contexts, supporting different requirements of indexation and retrieval process on several areas; and (iv) *shadow* internal structure follows interoperability standards and can be associated into the semantic Web scenario.

In conclusion, this work proposes a different approach to handle document formats diversity in the annotation context. The SDA approach uses a unaware file type reading to generate a interoperable descriptor that 'explodes' a document into addressable fragments. Once the descriptor is produced, present annotation standards can be used on a Web-based annotation process.

Referências

- [1] H. Cunningham. GATE, a general architecture for text engineering. *Computers and the Humanities*, 36(2):223–254, 2002.
- [2] T. Hey, S. Tansley, and K. Tolle, editors. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, Washington, 2009.
- [3] J. Kahan. Annotea: an open RDF infrastructure for shared Web annotations. *Computer Networks*, 39(5):589–608, Aug. 2002.
- [4] M. Koivunen, R. Swick, and E. Prud'hommeaux. Annotea shared bookmarks. In *Proc. of the KCAP 2003 workshop on knowledge markup and semantic annotation*, pages 25–26. Citeseer, 2003.
- [5] A. Santanchè, M. Mota, D. P. Costa, N. Oliveira, and C. O. Dalforno. Componere – web authoring based on components. In *Proc. of XV Brazilian Symp. on Multimedia and the Web*, 2009.

22 Empacotamento Tridimensional e Programação por Restrições

Autores: Pedro Hokama⁵ e Flávio Keidi Miyazawa

Consideramos um problema que combina os problemas do Caixeiro Viajante (TSP) e do Carregamento em Contêiner (3opp), que chamamos por Problema do Caixeiro Viajante com Restrições de Empacotamento Tridimensional (3D-TSP-L). Neste problema, um veículo deve partir carregado de uma origem e entregar caixas em pontos pré-definidos para seus clientes. Cada cliente tem um conjunto de caixas que deve receber e o objetivo é minimizar o custo de deslocamento do veículo. As caixas devem ser retiradas a partir do fundo do contêiner do veículo e a remoção das caixas de um cliente não pode ser obstruída pelas caixas a serem descarregadas posteriormente. Assim, diferentes rotas exigem a geração de diferentes configurações de empacotamento.

Hokama, Azevedo e Miyazawa [2] mostram um algoritmo *branch-and-cut* para resolver o problema, no caso particular onde deseja-se encontrar um empacotamento que possa ser realizado por uma mão mecânica, onde uma caixa só pode ser empacotada na frente, a direita, ou acima de outra. Esse estudo preliminar mostrou que o algoritmo gasta a maior parte do tempo para realizar o empacotamento e, portanto, é importante otimizar essa rotina.

Abordagem: Para resolver o problema do empacotamento com ordem estamos apresentando uma abordagem baseada em programação por restrições (CP), que na última década tem sido considerada uma promissora ferramenta para problemas onde uma formulação em programação linear inteira é complicada. Uma formulação em CP exige a identificação de variáveis, seus domínios, e de um conjunto de restrições que devem ser satisfeitas.

Uma modelagem envolvendo programação por restrições para resolver o problema de empacotamento 3opp, foi apresentada por Martello, Pisinger, Vigo, Boef e Korst em [3]. Nesse trabalho os autores consideram a posição relativa entre as caixas como variáveis e {esquerda, direita, acima, abaixo, afrente, atrás} como domínio.

Nossa abordagem utiliza as posições (x_i, y_i, z_i) de cada caixa i como variáveis, e o domínio de cada uma como os possíveis pontos de discretização onde uma caixa pode ser alocada. Um ponto de discretização da largura, é um valor $i \leq W$ (largura do contêiner) que pode ser obtido por qualquer combinação das larguras dos itens. Denotamos esse conjunto de pontos como P, e respectivamente fazemos o mesmo para a altura e a profundidade, obtendo os conjuntos Q e R.

Segundo Herz [1], qualquer solução viável para o problema possui uma solução correspondente nos pontos de discretização, dessa forma não perdemos nenhuma solução ao limitar o domínio a esses pontos.

Um domínio mais restrito, porém que elimina algumas soluções, são os conjuntos de pontos formados pelos chamados *raster points* descritos por Scheithauer [4].

Dos conjuntos P, Q e R, podemos obter os conjuntos de *raster points* \tilde{P} , \tilde{Q} e \tilde{R} , respectivamente, da seguinte forma: $\tilde{P} := \{\langle W - p \rangle_P | p \in P\}$; $\tilde{Q} := \{\langle H - q \rangle_Q | q \in Q\}$; $\tilde{R} := \{\langle D - r \rangle_R | r \in R\}$; onde $\langle s \rangle_A = \max\{a \in A | a \leq s\}$;

⁵Esta pesquisa é financiada pela FAPESP (Processo 2009/13270-0).

Tabela 1: Tempo Computacional

Class	He	He	He	He	He	He	Ho	Ho	Ho	Ho	Ho	Ho	Ho	Ho	Ho	Ho	Ho	Ho
Vol0	50	50	64	64	64	64	51	51	51	51	52	52	52	52	61	61	61	61
itens	21	21	25	25	25	25	25	25	25	25	21	21	21	21	29	29	29	29
clientes	10	21	02	06	12	25	02	06	12	25	02	05	10	21	02	07	14	29
DP	8	11	84	107	157	276	4	21	18	20	1	6	9	5	83	110	445	322
RP	9	20	77	98	280	171	1	13	7	11	0	2	2	2	21	57	40	113
Robot	160	328	-	-	567	-	0	1	0	0	0	0	0	0	-	-	-	-

Resultados: Os algoritmos descritos foram implementados em C, utilizando o IBM ILOG CPLEX CP Optimizer como resolvidor de Programação por Restrições. Os testes foram executados em um único core de um Intel Core 2 Quad 2.40GHz. O tempo máximo permitido para cada instância foi de 3600 segundos. A tabela 1 mostra o tempo em segundos que foi necessário para encontrar a solução de cada instância. Onde não foi possível encontrar uma solução, ou afirmar que essa não existia dentro do limite do tempo, o espaço na tabela indica “-”. A linha **Vol** indica o volume aproximado dos itens em relação ao contêiner, a linha **DP** mostra o tempo utilizando os pontos de discretização, e a linha **RP** mostra o tempo utilizando os *raster points*. E, por fim, a última linha mostra o tempo utilizando a rotina de [2]. As instâncias da classe **Ho** possuem itens repetidos, enquanto a classe **He** possui em geral todos os itens diferentes.

Conclusões: Nesse trabalho estamos apresentando uma nova abordagem para resolver o problema do 3D-TSP-L, em particular o subproblema 3opp. Os resultados preliminares são bastantes promissores, conseguindo resolver em tempo computacionalmente viável instâncias de pequeno e médio porte.

Referências

- [1] J. C. Herz. Recursive computational procedure for two-dimensional stock cutting. *IBM J. Res. Dev.*, 16:462–469, September 1972.
- [2] P. Hokama, B. L. P. de Azevedo, and F. K. Miyazawa. Problema do caixeiro viajante tridimensional com restrição de ordem. *XXXV Conferencia Latinoamericana de Informática*, 2009.
- [3] S. Martello, D. Pisinger, D. Vigo, E. D. Boef, and J. Korst. Algorithm 864: General and robot-packable variants of the three-dimensional bin packing problem. *ACM Trans. Math. Softw.*, 33, March 2007.
- [4] G. Scheithauer. Equivalence and dominance for problems of optimal packing of rectangles. In *Ricerca Operativa*, volume 27, pages 3–34. 1997.

23 Simulação de multidões e planejamento probabilístico para otimização de intersecções controladas por semáforo

Autores: Renato Schattan P. Coelho, Siome Klein Goldenstein e Jacques Wainer

Há muito tempo estudam-se maneiras de simular o comportamento do tráfego de carros, devido ao grande custo econômico e ambiental dos engarrafamentos [3]. Apenas na cidade de São Paulo calcula-se que o gasto anual com trânsito seja da ordem de R\$ 33 bilhões⁶. Embora muitos estudos tenham se dedicado a modelar o tráfego poucos usam estes modelos para a automação da tomada de decisões.

Avanços no sentido da automação de tomada de decisões podem ser obtidos através de algoritmos de planejamento probabilístico [4]. Um bom exemplo de seu uso para controle de ambientes são os resultados obtidos em Pesquisa Operacional [1]. No entanto, uma das maiores dificuldades para o uso eficiente dos planejadores já desenvolvidos reside na descrição de problemas; geralmente, são utilizados problemas sem interesse prático ou é necessário que um especialista descreva o problema de planejamento probabilístico [5].

Nosso projeto tem como principal objetivo otimizar o controle de semáforos para maximizar o fluxo de veículos em uma interseção ou em uma rede de ruas. Nossa proposta para alcançar este objetivo é a criação de um sistema que combine a simulação de multidões com algoritmos de planejamento probabilístico.

O sistema proposto é composto por quatro módulos: uma interface para a descrição das redes de ruas que devem ser estudadas, um simulador de multidões, um planejador probabilístico, que usará os dados gerados pelo simulador, e um módulo para a integração dos três outros módulos.

Ao longo do projeto estudamos vários modelos utilizados para simulação de multidões tendo em vista algumas características que melhor se adequam ao nosso projeto, como: flexibilidade, objetivos diferentes para cada agente e capacidade de modelar obstáculos com efeito limitado.

O modelo de simulação de multidões escolhido para o sistema foi o *Cellular Automata* [2]. Neste modelo o espaço e o tempo são discretizados. O ambiente evolui de acordo com regras de atualização das células que levam em conta apenas um pequeno número de células vizinhas. Com estas características o modelo consegue uma boa fidelidade e uma grande velocidade de simulação.

Para o planejamento probabilístico escolhemos o modelo dos Processos Markovianos de Decisão (*Markov Decision Processes* - MDP) [1], dado que este possui propriedades muito interessantes para modelar o problema que queremos resolver, como a imprecisão nos efeitos das ações e a independência do histórico no efeito das ações. Um MDP pode ser descrito através de suas componentes: um espaço de estados S finito e não vazio, um conjunto de ações A , uma função de transição de estados $F(s, a)$, uma distribuição de probabilidades $P(\cdot|s, a)$ sobre $F(s, a)$ e uma função valor $V(h)$ que associa um valor a um histórico de estados e ações [4].

O primeiro passo para a integração destas duas técnicas é definir quais são as variáveis

⁶Dados divulgados no segundo Seminário nacional de gestão de frotas, pelo vice-presidente da Fundação Getúlio Vargas, Marcos Cintra Cavalcanti de Albuquerque.

do mundo que definem os estados (número de carros em cada rua, se a próxima quadra pode receber mais carros, quantidade de carros passando pelo cruzamento) para associar situações no modelo a um estado no problema de planejamento probabilístico. Como o modelo de planejamento probabilístico não leva em conta tantas variáveis quanto o modelo de simulação cada estado do planejador representa diferentes estados do simulador. Ou seja, os estados considerados pelo planejador serão arquétipos dos estados encontrados no simulador. É neste momento que surge a imprecisão no efeito das ações.

Esta maneira de unir os algoritmos das duas áreas apresenta algumas características interessantes, como a facilidade de se trocar as partes do sistema independentemente (interface, simulador e planejador) e a possibilidade de reaproveitar os cálculos ao obter novas distribuições de probabilidade.

Atualmente o trabalho se encontra na fase final do desenvolvimento do simulador de multidões. Além disso, desenvolvemos a interface gráfica para a descrição das ruas, facilitando o desenvolvimento de casos de teste e o uso do programa para o usuário final.

Para analisar se a política devolvida pelo sistema efetivamente melhora o fluxo de veículos no sistema estudado, faremos comparações entre a política devolvida e a que se utiliza nos sistemas atuais de semáforos e com as políticas devolvidas por outros algoritmos de otimização de fluxo.

Referências

- [1] C. Boutilier, T. Dean e S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.
- [2] D. Chowdhury, L. Santen e A. Schadschneider. Statistical physics of vehicular traffic and some related systems. *Physics Reports*, 329:199, 2000.
- [3] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4):1067–1141, December 2001.
- [4] S. Russell e P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2 ed., December 2002.
- [5] F. Trevizan, F. Cozman e L. Barros. Planning under risk and knightian uncertainty. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2007.

24 Visualization Techniques for Content-Based Image Retrieval

Authors: Sheila M. Pinto-Cáceres and Ricardo da S. Torres

Recent advances in digital technologies, new image acquisition devices and the quick growth of Internet have enabled the existence of huge image collections. These vast data sets need to be efficiently organized and managed in order to be retrieved and consequently used. This scenario evidence the necessity of a friendly way to access and browse the large image universe based on the user interest minimizing the display of non-relevant images.

A CBIR system offers mechanisms needed to search and retrieve images based on visual properties such as color, texture, shape, etc. Extensive work has been done in CBIR context. However, most of existing CBIR systems have employed a too rigid layout to arrange the results, restricting users to explore the result image universe using a list or grid based distribution with, in some cases, no possible manipulation over the retrieved set. Moreover, techniques for visualizing images in CBIR systems have been poorly explored [1].

This work presents a comparative study of several visualization techniques. As a result of this study, we have distinguish three main groups of visualization methods. Traditional approaches use a list or a grid structure to display the images and is used by most of the existing CBIR systems. Radial methods display the result set in circular or elliptical forms [1]. Clustering methods exploit clustering algorithms in order to analyze the similarities between all the search results displaying similar elements of the result set close to each other. On one hand, traditional and radial structures offer a good navigability and exploration of the result set in an organized way. However, they generally do not respect the intrinsic relationship among the results, displaying different similarity degrees at the same physical distance to the query pattern. On the other hand, cluster-based techniques provide a better understanding of the universe of available results. This approach usually focuses on a large amount of information without, in many cases, taking care of an adequate design to distribute the clusters, which entails overlapping and confusion for the users.

Different from all previous approaches, in this work, we present a novel approach for the interactive search that displays the result set in a more flexible and intuitive manner. The proposed strategy combines the advantages of those strategies into a single method, named Clustering Set as illustrated in Figure 4. Our approach groups together elements with similar content into clusters, which are displayed in a radial manner as a well-defined tree. In order to maintain the flexibility, we do not restrict the proposed scheme to use a specific clustering algorithm. Then, the query pattern is placed in the the center of the visualization display. The clusters are circularly distributed around the query in a clockwise order of similarity, which is represented by visual marks. Same strategy is also applied inside each of the clusters in order to provide a more intuitive understanding of the display.

Moreover, the proposed approach enables a query refinement by using hierarquical clustering. It also avoids overlapping in most cases which represents a valuable advantage in the visualization area.

We have developed a framework to support the studied and proposed methods. In this framework, a visual element can be an image or a visualization structure and we can

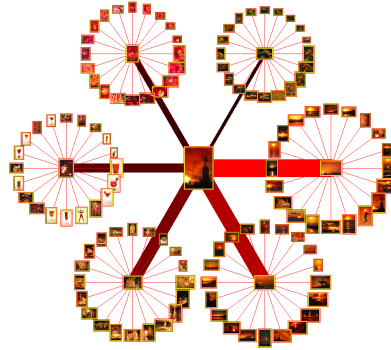


Figura 4: Abordagem proposta.

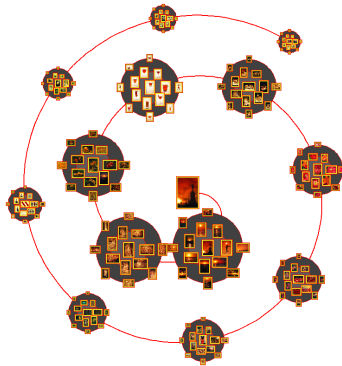


Figura 5: Flexibilidade do framework.

combine elements, so it allows the dynamic creation of fractal and hybrid structures. By using this framework, we have created new hybrid strategies by combining previous ones. Figure 5 shows an hybrid example of the framework flexibility. This framework also provides a friendly interface permitting the customization of several visualization parameters in order to let the user extract interest information.

Results from a subjective evaluation with 38 users confirms that the proposed strategy clearly outperforms other visualization strategies by considering several criteria such as user satisfaction, understanding degree, easiness of finding some elements, etc. The results also show that our strategy makes the navigation more coherent and engaging to users. We are convinced that such a visualization technique is a valuable contribution for image browsing and retrieval systems.

We have also extended this work for video retrieval through video stories [2] and we have explored another type of navigation using Divisive-Agglomerative Hierarchical Clustering trees [3].

Referências

- [1] da S. Torres, R., Silva, C. G., Medeiros, C. B., and Rocha, H. V. (2003). Visual structures for image browsing. *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management USA* pp. 49–55.
- [2] Pinto-Cáceres, S., Almeida, J., Neris, V., Baranauskas, M., Leite, N., and da S. Torres, R. (2011). Navigating through Video Stories using Clustering Sets. *International Journal of Multimedia Data Engineering and Management*,.
- [3] Almeida, J., Pinto-Cáceres, S., da S. Torres, R., and Leite, N. (2011). Intuitive Video Browsing Along Hierarchical Trees. Technical report, Institute of Computing - State University of Campinas.

25 Rearranjos em Genomas

Autores: Priscila do Nascimento Biller e João Meidanis

Definimos como filogenia a história evolutiva de um organismo, que descreve os vários organismos de que aquele descende, juntamente com a sua relação com outros organismos vivos. Uma forma de representar a filogenia são as árvores filogenéticas, onde os organismos modernos ficam nas folhas, os ancestrais nos nós internos e as arestas indicam relacionamentos evolucionários.

Já os organismos podem ser representados por seu genoma, que é a sequência de DNA completa de um conjunto de cromossomos. Ele contém toda a informação necessária para codificar as proteínas, que realizam diversas funções celulares.

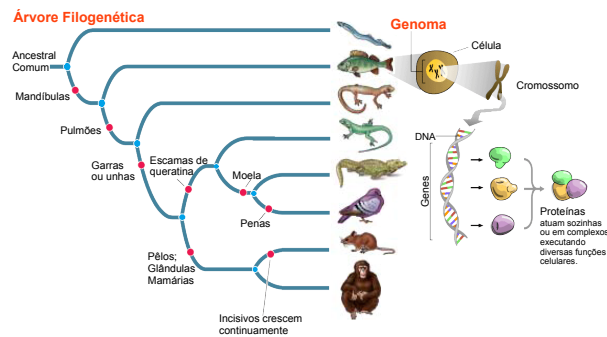


Figura 6: Exemplo de representação da filogenia e do organismo. [1]

Compreender os mecanismos através dos quais os organismos evoluem tem diversas aplicações. Além de auxiliar na classificação e identificação de espécies, também ajuda a mapear quais partes do genoma estão associadas a certas características ou sintomas desenvolvidos no indivíduo. Na área de saúde humana e animal, por exemplo, é interessante conhecer quais partes do genoma desencadeiam determinada doença e em quais situações. No agronegócio, é interessante conhecer quais partes do genoma estão relacionadas com uma determinada característica desejável.

Computacionalmente, existem diversos métodos que abordam o problema da inferência filogenética. Um deles é o método da parcimônia, onde o objetivo é encontrar a árvore filogenética mais parcimoniosa, isto é, a árvore que requer o menor número de passos evolucionários (custo) para explicar o conjunto de dados observados. O Problema da Parcimônia possui duas variantes, definidas a seguir.

No **Problema Pequeno da Parcimônia**, dada a topologia de uma árvore T , um conjunto de genomas G (folhas de T) e uma operação de rearranjo R , que determina quais eventos evolutivos são permitidos, o objetivo é construir um conjunto de genomas H e associá-los aos nós internos de T , minimizando o custo total da árvore.

No **Problema Grande da Parcimônia**, a topologia da árvore não é conhecida: são dados apenas um conjunto de genomas G (que estão nas folhas de T) e uma operação de rearranjo R . O objetivo é reconstruir uma árvore filogenética T , tal que o custo total de T seja mínimo.

A complexidade de ambos os problemas varia conforme a operação de rearranjo. Re-

centemente, em 2009, Feijão e Meidanis [2] propuseram uma nova operação de rearranjo: o *single-cut-or-join* (SCJ). Esta operação modela diversos eventos evolutivos tradicionais, como reversão, transposição, translocação, fissão e fusão. A vantagem do SCJ em relação a outras operações de rearranjo é que ele resolve de forma polinomial diversos problemas, entre eles, o Problema Pequeno da Parcimônia.

Por ser uma proposta recente, ainda não existem estudos que verifiquem sua eficiência na prática. Dessa forma, a proposta desse trabalho é verificar, com dados simulados, em que medida SCJ é capaz de reconstruir histórias evolutivas a partir dos genomas dados. Em um segundo momento, com dados reais, verificaremos como os resultados de reconstrução de histórias evolutivas por SCJ comparam-se às árvores mais aceitas pela comunidade científica.

No Problema Pequeno da Parcimônia, verificaremos a qualidade dos genomas ancestrais inferidos através do algoritmo de Fitch [3], que é a abordagem que o SCJ utiliza para resolver de forma exata e em tempo linear este problema. Já no Problema Grande da Parcimônia verificaremos a qualidade da topologia da árvore inferida em relação à topologia da árvore real. Como este segundo problema é NP-difícil, implementamos a heurística de inclusão passo-a-passo e também um algoritmo *branch-and-bound*. Abaixo temos alguns resultados da aplicação do SCJ no Problema Grande da Parcimônia utilizando a heurística de inclusão passo-a-passo:

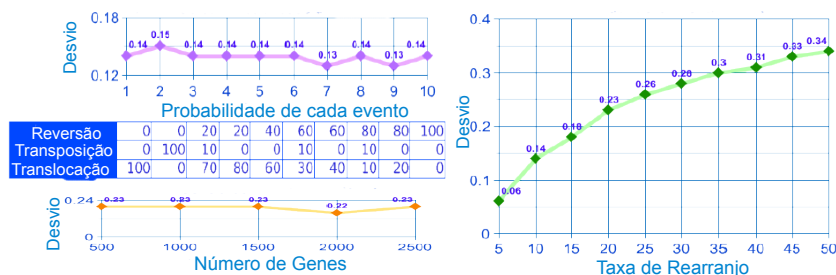


Figura 7: Impacto de diversos fatores na operação SCJ.

Os resultados preliminares mostram que a inferência filogenética por SCJ não é influenciada por fatores como tamanho do genoma, número de genomas e probabilidade de cada evento. Essa estabilidade torna a operação SCJ bastante promissora e estão sendo realizadas outras análises para comprovar a eficiência da operação.

Referências

- [1] B. Purves, D. Sadava, G. H. Orians, and H. C. Helles. *Life: The Science of Biology*. Sinauer Associates and W. H. Freeman, 2003.
- [2] P. Feijão and J. Meidanis. SCJ: a variant of breakpoint distance for which sorting, genome median and genome halving problems are easy. *Algorithms in Bioinformatics*, 5724:85-86, 2009.
- [3] W. M. FITCH. Toward defining the course of evolution: Minimum change for a specific tree topology. *Systematic zoology*, 20:406-416, 1971.

26 Aplicação de Técnicas de Visão Computacional e Aprendizado de Máquina para a Detecção de Exsudatos Duros em Imagens de Fundo de Olho

Autores: Tiago Carvalho, Siome Goldenstein, Jacques Wainer

A *diabetes mellitus* (DM) vem crescendo de forma rápida e segundo dados fornecidos pela Organização Mundial de Saúde (OMS) 300 milhões de pessoas desenvolverão a doença até 2025 [1]. A retinopatia diabética (RD) é um dos tipos de complicações provocados pela DM no organismo.

Pacientes portadores de RD tendem a desenvolver diversos tipos de lesões em sua retina. Uma dessas lesões denomina-se exsudatos duros (ED) e é provocada pelo vazamento de gordura através dos capilares da retina.

Neste trabalho, desenvolvemos um método robusto e com alta taxa de sensibilidade para detecção de ED em imagens de fundo de olho. Nosso método é baseado em pontos de interesse e dicionários visuais o que o torna de fácil extensão para ser aplicado em outros tipos de anomalias.

Ao se encontrar pontos de interesse (PoI) e se caracterizar a região ao redor de cada um desses pontos obtemos uma forma de representar o conteúdo visual de uma imagem. É desejável escolher pontos de interesse invariantes a escala e orientação para que essa representação seja robusta a algumas transformações da imagem. Para isso utilizamos o método denominado Speeded-Up Robust Features (SURF) [2].

Para preservar o poder de distinção dos PoI e aumentar sua generalização, utilizamos o conceito de dicionários visuais (DV) [3]. Um dicionário de tamanho x é baseado nos $\frac{x}{2}$ PoI mais representativos de imagens normais, e pelos $\frac{x}{2}$ PoI mais representativos de imagens com anomalia. Tais pontos são escolhidos com base nas imagens de treinamento e denominam-se palavras visuais.

Descritas as imagens em função de seus PoI e do DV, utilizamos o *Support Vector Machines* (SVM) [4] para realizar a classificação. Neste processo a base de dados é dividida em dois conjuntos: treino e teste. Os parâmetros de ajuste do classificador são obtidos utilizando um processo de busca automática sobre o conjunto de treino. Para validar os resultados, utilizamos o processo de validação cruzada baseada em cinco dobras.

Uma vez classificadas as imagens, expressamos nossos resultados utilizando medidas de probabilidade denominadas sensibilidade (SE) e especificidade (ES) as quais indicam respectivamente a taxa de verdadeiros positivos e a taxa de falsos positivos.

Como resultado, nosso método obteve utilizando um dicionário com 500 palavras visuais uma taxa de 90% de sensibilidade com 87% de especificidade. A taxa de erro igual foi de 88% para 500 palavras e a área sob a curva de 95.3%.

Assim, podemos concluir que nosso método baseado em pontos de interesse e dicionários visuais, possui um poder descritivo adequado ao problema de detecção de exsudatos em imagens de fundo de olho. Além disso, novos experimentos apontam que o método também mostra-se eficaz quando utilizado para detectar outros tipos de anomalia provenientes da retinopatia diabética.

Referências

- [1] K. W. Tobin, E. Chaum, V. P. Govindasamy, and T. P. Karnowski. Detection of Anatomic Structures in Human Retinal Imagery *IEEE Transactions on Medical Imaging*, 26(12):1729–1739, 2007.
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (SURF). *Computer Vision Image Understand*, 110(3):346–359, 2008.
- [3] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. *In 9th IEEE International Conference on Computer Vision*, pages 1470–1477, 2003.
- [4] K. P. Bennett and C. Campbell. Support vector machines: Hype or hallelujah? *Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining*, 2(2):1–13, 2000.

27 Programação do WTD

Nesta Seção apresentamos a programação do VI Workshop de Teses, Dissertações e Trabalhos de Iniciação Científica, do Instituto de Computação (IC) da UNICAMP. As apresentações foram divididas em duas salas (Veja as Tabelas 8 e 9) e avaliadas por diversos professores do instituto (Veja a Tabela 10).

VI WTD - INSTITUTO DE COMPUTAÇÃO - UNICAMP - 11/05/2011					
Horário		Nome	Orientador	Título	Local
8:00		Abertura		Abertura do VI WTD [Apresentação, Estatística, etc.]	
8:15	SI	M Vitor de Lima	Hélio Pedrini	Codificação fractal de vídeo	IC3.5 - 351
8:45	SI	M Hugo Augusto Alves	André Santanchè	Integrating and Enriching Folksonomies and Ontologies	IC3.5 - 351
9:15	TC	M Igor Ribeiro de Assis	Cid Carvalho de Souza	O Problema do Recorte com Custo nas Conversões	IC3.5 - 351
9:45	TC	M Lucas de Oliveira	Cid Carvalho de Souza	Problema do Corredor de Comprimento Mínimo: Algoritmos exato, aproximativos e heurísticas	IC3.5 - 351
10:15		Apresentação de Pôsters	Sala de Coffee-Break		
10:45	SI	M César Christian Castelo Fernández	Pedro J. Rezende e Alexandre X. Falcão	Novos Algoritmos de Aprendizado para Classificação de Padrões Utilizando Floresta de Caminhos Ótimos	IC3.5 - 351
11:15	TC	D Evandro Cesar Bracht	Flávio Keidi Miyazawa	Problemas de Empacotamento com restrições práticas	IC3.5 - 351
11:45	TC	M Pedro Hokama	Flávio Keidi Miyazawa	Empacotamento Tridimensional e Programação por Restrições	IC3.5 - 351
12:15		Intervalo			
14:00	TC	D Carlos Eduardo de Andrade	Flávio Keidi Miyazawa	Leilões e Algoritmos	IC3.5 - 351
14:30	TC	D Diego de Freitas Aranha	Julio López	Implementação eficiente de algoritmos criptográficos em arquiteturas modernas	IC3.5 - 351
15:00	TC	M João Paulo Pereira Zanetti	João Meidanis	Complexidade de Construção de Árvores PQR	IC3.5 - 351
15:30	SI	D Danillo Roberto Pereira	Jorge Stolfi	FITTING 3D DEFORMABLE MODELS TO MICROSCOPE IMAGES	IC3.5 - 351
16:00		Apresentação de Pôsters	Sala de Coffee-Break		
16:45	TC	D Jefferson Luiz Moisés da Silveira	Eduardo Candido Xavier	Algoritmos Aproximados para Problemas de Empacotamento em Faixa com Restrição de Ordem	IC3.5 - 351
17:15	TC	M Thiago de Paulo Faleiros	Eduardo Candido Xavier	Algoritmos para o problema de Particionamento	IC3.5 - 351
17:45	SC	D Gabriel Dieterich Cavalcante	Paulo Lício de Geus	Utilizando visualização para fins de segurança	IC3.5 - 351
18:15		M Gustavo Prado Alkmim	Nelson Fonseca	Mapeamento de Redes Virtuais em Substratos de Rede	IC3.5 - 351

Figura 8: Programação das apresentações na sala 351 do IC-3,5.

VI WTD - INSTITUTO DE COMPUTAÇÃO - UNICAMP						
8:00	Abertura			Abertura do VI WTD [Apresentação, Estatística, etc.]		Sala
8:15	SI	M	Sheila Maricela Pinto Cáceres	Ricardo Torres	Técnicas de Visualização para Sistemas de Recuperação de Im	IC3 - 316
8:45	SI	D	Daniel Carlos Guimarães Pedronette	Ricardo Torres	Using Context Information for Data Fusion in Content-Based Im	IC3 - 316
9:15	SI	D	Joana Esther Gonzales Malaverri	Claudia Medeiros	Managing Data Provenance for Biodiversity Information	IC3 - 316
9:45	SI	D	Ivo Kenji Koga	Claudia Medeiros	Gerenciamento de dados de sensores para Sistemas de Informação de Biodiversidade	IC3 - 316
10:15	Apresentação de Pôsters			Sala de Coffee-Break		
10:45	SI	M	Vitor Ciaramella	Mario Cortes	Jogo simulador para aprendizagem em gestão de projetos	IC3 - 316
11:15	SI	M	Marcelo Invert Palma Salas	Eliane Martins	METODOLOGIA DE TESTES DE ROBUSTEZ PARA SEGURANÇA EM SERVIÇOS WEB	IC3 - 316
11:45	SI	D	Juliana Galvani Gregghi	Eliane Martins	Geração Automática de Casos de Uso a partir de Especificação em Língua Natural	IC3 - 316
12:15	Intervalo					
14:00	SI	M	Alan Braz	Cecilia Rubira	Método Ágil aplicado ao Desenvolvimento de Software Confiável baseado em Componentes	IC3 - 316
14:30	SI	D	Heiko Hornung	M. Cecília C. Baranauskas	Interaction Design in the Pragmatic Web (working title)	IC3 - 316
15:00	SI	D	Ricardo Edgard Caceffo	Heloisa Vieira da Rocha	Arquitetura para Suporte a Aplicações Ubíquas que Viabilizam e	IC3 - 316
15:30	SI	D	André Constantino da Silva	Heloísa Vieira da Rocha	Ambiente de EaD com Interação Multimodal: uma proposta para aumentar a acessibilidade e educabilidade em ambientes de Educação a Distância	IC3 - 316
16:00	Apresentação de Pôsters			Sala de Coffee-Break		
16:45	SI	M	Ivelize Rocha Bernardo	André Santanchè	Explicitação de Esquema Orientada ao Contexto para Promover Interoperabilidade Semântica	IC3 - 316
17:15	SI	D	Gizelle Sandrini Lemos	Eliane Martins	Avaliação de resultados de teste através do uso de algoritmos dealinhamento de sequências	IC3 - 316
17:45	SI	D	Thaise Yano	Eliane Martins	Uma abordagem de otimização multiobjetiva em teste baseado em modelo	IC3 - 316
18:15	SI	D	Thiago Vallin Spina	Alexandre Xavier Falcão	Segmentação interativa de objetos em imagens e vídeos baseada em grafos e modelos nebulosos de conhecimento de conteúdo	IC3 - 316
18:45	Encerramento					

Figura 9: Programação das apresentações na sala 316 do IC-3.

Professor	Horário	Local	Professor	Horário	Local
Cid	08:00 - 10:15	IC3.5 - 351	Borin	08:00 - 10:15	IC3 - 316
Christiane	08:00 - 10:15	IC3.5 - 351	Dahab	08:00 - 10:15	IC3 - 316
Hélio	08:00 - 10:15	IC3.5 - 351	Guido	10:45 - 12:15	IC3 - 316
Flávio	10:45 - 12:15	IC3.5 - 351	Paulo Licio	10:45 - 12:15	IC3 - 316
Orlando	10:45 - 12:15	IC3.5 - 351	Rubira	14:00 - 16:00	IC3 - 316
Neucimar	10:45 - 12:15	IC3.5 - 351	Beatriz	14:00 - 16:00	IC3 - 316
Guilherme	14:00 - 16:00	IC3.5 - 351	Eliane	16:45 - 18:45	IC3 - 316
Eduardo Xavier	14:00 - 16:00	IC3.5 - 351	Buzato	16:45 - 18:45	IC3 - 316
Anderson	16:45 - 18:45	IC3.5 - 351			

Figura 10: Professores das bancas de avaliação.