

Universidade Federal de Uberlândia
Faculdade de Computação

VI FACOM TECHWEEK & XIII WORKSHOP DE TESES E DISSERTAÇÕES EM CIÊNCIA DA COMPUTAÇÃO

Anais

10 a 14 de junho de 2019

ISSN: 2447-0406



Uberlândia
2019

VI FACOM Techweek

Anais do XIII WTDC

Uberlândia, MG, Brasil

10 à 14 de junho de 2019

Comissão Organizadora

Christiane Regina Soares Brasil (UFU)
Fernanda Maria da Cunha Santos (UFU)
Luiz Gustavo Almeida Martins (UFU)
Murillo Guimarães Carneiro (UFU)
Paulo Henrique Ribeiro Gabriel (UFU)
Renan Gonçalves Cattelan (UFU)
Wendel Alexandre Xavier de Melo (UFU)

Organização, Execução, Promoção e Realização

Universidade Federal de Uberlândia (UFU)
Faculdade de Computação (FACOM/UFU)

Apoio

Algar Telecom
Beyond Soluções
Sankhya Gestão de Negócios
TQI

FACOM Techweek e WTDCC

A TechWeek é um evento realizado pela Faculdade de Computação (FACOM) da Universidade Federal de Uberlândia (UFU), que busca promover atualização técnica e apresentar as últimas tendências tecnológicas, reunindo e proporcionando uma maior interação entre estudantes, professores, profissionais e empresas das áreas de Tecnologia da Informação. Trata-se de um evento tradicional, já em sua sexta edição, composto por palestras, minicursos, mesas redondas, competições técnicas e apresentação de trabalhos e pesquisas científicas.

Realizado em conjunto com a Techweek desde 2016, o Workshop de Teses e Dissertações em Ciência da Computação (WTDCC) é um evento do Programa de Pós-Graduação em Computação (PPGCO) da UFU, na sua décima terceira edição, o qual conta com a participação de estudantes de graduação e pós-graduação de universidades de Uberlândia e região. Dentre os objetivos do WTDCC destacam-se: incentivar a troca de experiência entre pesquisadores, apoiar os trabalhos produzidos pelos discentes, promover a interação entre os diversos grupos de pesquisa e refletir sobre a atual situação deles no cenário nacional. Nesse sentido, o evento consiste em um importante espaço de integração para a troca de experiências acadêmico-científicas, objetivando o desenvolvimento da ciência e da tecnologia nesse domínio do conhecimento.

Prefácio

O Workshop de Teses e Dissertações em Ciência da Computação foi realizado junto com a VI FACOM Techweek e teve como objetivo divulgar os projetos de pesquisa científica e tecnológica na área de computação, realizados por estudantes de graduação e pós-graduação da Universidade Federal de Uberlândia (UFU) e região. Também teve por objetivo contribuir para a formação dos participantes despertando o interesse pelas descobertas científicas e pela resolução de problemas complexos.

Na sua décima terceira edição, o WTDCD contou com palestras temáticas e apresentação de trabalhos. As palestras compreenderam tópicos relacionados às linhas de pesquisa desenvolvidas pelo Programa de Pós-Graduação em Ciência da Computação (PPGCO/UFU), bem como sobre os desafios da pós-graduação e a situação atual do programa em nível nacional e internacional. As apresentações de trabalho, em formato de pôster, contemplaram três trilhas relacionadas à formação dos discentes: trabalhos de graduação, mestrado e doutorado, sendo que os melhores trabalhos em cada categoria foram premiados.

Além de contar com os trabalhos de dezenas de autores, o WTDCD também contou com boa presença do público, interessado em conhecer mais sobre os trabalhos desenvolvidos em Uberlândia e região. Nesse sentido, agradecemos a todos os alunos e orientadores que submeteram seus trabalhos para o evento e também ao público que prestigiou a apresentação de tais trabalhos.

Por fim, agradecemos à disponibilidade e empenho dos docentes que colaboraram na revisão e avaliação dos trabalhos e também às empresas e instituições que nos proveram todo apoio necessário para realizar este evento.

Prof. Luiz Gustavo Almeida Martins
Prof. Murillo Guimarães Carneiro
Coordenadores do XIII WTDCD

Organização

Coordenação Geral

Prof. Renan Gonçalves Cattelan (UFU)

Coordenação do XIII WTDCC

Prof. Luiz Gustavo Almeida Martins (UFU)

Prof. Murillo Guimarães Carneiro (UFU)

Comissão Organizadora

Profa. Christiane Regina Soares Brasil (UFU)

Profa. Fernanda Maria da Cunha Santos (UFU)

Prof. Luiz Gustavo Almeida Martins (UFU)

Prof. Murillo Guimarães Carneiro (UFU)

Prof. Renan Gonçalves Cattelan (UFU)

Prof. Paulo Henrique Ribeiro Gabriel (UFU)

Prof. Wendel Alexandre Xavier de Melo (UFU)

Revisores do XIII WTDCC

Prof. Alexsandro Santos Soares (UFU)

Prof. André Ricardo Backes (UFU)

Prof. Bruno Augusto Nassif Travençolo (UFU)

Prof. Carlos Roberto Lopes (UFU)

Prof. Christiane Regina Soares Brasil (UFU)

Prof. Dino Rogério Coinete Franklin (UFU)

Profa. Elaine Ribeiro de Faria Paiva (UFU)

Prof. Ivan da Silva Sendin (UFU)

Prof. Marcelo de Almeida Maia (UFU)

Prof. Marcelo Keese Albertini (UFU)

Prof. Marcelo Zanchetta do Nascimento (UFU)

Profa. Maria Adriana Vidigal de Lima (UFU)

Profa. Maria Camila Nardini Barioni (UFU)

Prof. Paulo Henrique Ribeiro Gabriel (UFU)

Prof. Rodrigo Sanches Miani (UFU)

Prof. Renan Gonçalves Cattelan (UFU)

Prof. Rivalino Matias Júnior (UFU)

Prof. Wendel Alexandre Xavier de Melo (UFU)

Equipe de Trabalho

Antonio Carlos Neto

Breno Corrêa Silva Costa

Bruna Alves Oliveira

Bruno de Paula Nery
Fabrício Fernandes Ziliotti
Felipe Augusto Ferreira de Castro
Gabriel Rocha Gomes
Gabriel Valentin Tiburcio
Giullia Rodrigues de Menezes
Guilherme Henrique de Araújo Santos
Gustavo Silva de Oliveira
Higor Emanuel Souza Silva
Higor Raphael Faria e Sousa
Isaque dos Reis
Laís Saloum Deghaide
Lara Mondini Martins
Lucas Guimarães Mendes
Lucas Josino de Paula Gonçalves
Lucas Riccioppo Silva
Marcelo Mendonça Borges
Matheus Moreira de Camargo
Salomão Oliveira Alves
Vitor Manoel Gonçalves Teixeira
Vitor Martins Basso

Editoração dos Anais

Prof. Murillo Guimarães Carneiro (UFU)
Prof. Paulo Henrique Ribeiro Gabriel (UFU)

Sumário

I TRABALHOS DE GRADUAÇÃO	1
Análise de redes de espalhamento de fake news	2
Incorporação do local subcelular de proteínas e propensão a vacinas no software Pan- notator utilizando-se do software medpipe.	3
Agrupamento de dados via combinação de partições na presença de atributos faltantes	4
Construção de mecanismo para suportar a predição de tempos de resposta do Cas- sandra a partir de métricas da infraestrutura openstack	5
Análise da motivação dos alunos ingressantes em ciências da computação	6
Investigando entropia na classificação de tecidos histológicos de mama	7
Estudo de características urbanas por meio da análise de imagens por sensoriamento remoto	8
Aprendizado de máquina na avaliação de risco de crédito financeiro	9
Deteção de comunidades em redes complexas aplicado ao agrupamento de objetos invariantes	10
Análise de segurança em sistemas operacionais dedicados para internet of things (IOT)	11
Gerontecnologia: um estudo de caso para inclusão digital	12
V2C – protótipo de veículo conectado	13
Estudo comparativo de técnicas do estado da arte para predição de séries temporais não-estacionárias.	14
Estruturas de indexação baseadas em mapeamentos unidimensionais para otimização de consultas por similaridade	15
Análise do tráfego de rede gerado por veículos aéreos não tripulados comerciais . . .	16
Progressive web apps: características e aplicações	17
Uso de modelos probabilísticos para melhorar a eficiência da exploração de sequências de otimização do compilador	18
II TRABALHOS DE MESTRADO	19
Avaliação da capacidade de retenção de água em amostras de carne suína	20
Algoritmos evolutivos aplicados ao problema de escalonamento de tarefas em sistemas multiprocessados	21
Aprimorando o desempenho na cobrança de faltas de agentes jogadores de FIFA por meio de técnicas de detecção de objetos	22
Desenvolvimento de algoritmos e técnicas em Bioinformática para estudo de genomas de patógenos humanos procariotos	23
Predição de porções de proteínas que estimulam o sistema imunológico adaptativo utilizando uma rede neural convolucional (Deep Learning) programada em Lisp .	24
Classificação multirrotulo de alto nível por meio do uso de redes complexas	25
Descoberta antecipada de exploits usando redes sociais	26
Aprendizado de ranking de entidades aplicado aos dados do governo brasileiro	27
Orquestração osmótica de cloud slices	28
III TRABALHOS DE DOUTORADO	29
CryptoCeA: criptografia com autômatos celulares	30
Open-and-secure data framework (OSDFramework): a pragmatic proposal for building new applications from existing iot applications	31
Classificação de tumores de mama e colorretais baseada em percolação de imagens normalizadas	32
Geração de documentação a partir de consultas a serviços de busca	33
Learning coarse-grained code structures	34

Parte I

TRABALHOS DE GRADUAÇÃO

Análise de redes de espalhamento de fake news

Adriano Resende (Universidade Federal de Uberlândia)*; Daniel Abdala (Universidade Federal de Uberlândia)

adrianozeroleft@ufu.br; abdala@ufu.br

Resumo: O fenômeno das fake news não é recente, porém atualmente, suas consequências estão evidenciadas levando a casos de linchamento de pessoas e manipulação eleições. O presente estudo visa desenvolver uma ferramenta computacional para auxiliar o processo de identificação de fake news na rede social twitter. Usando as informações fornecidas pela API do twitter, data de publicação e perfis que retweetaram, será modelado um grafo dinâmico e pela análise desse grafo estima-se que seja possível classificar quais tweets são potencialmente dissimuladores de fake news.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Incorporação do local subcelular de proteínas e propensão a vacinas no software Pannotator utilizando-se do software medpipe.

Salomão Alves (Universidade Federal de Uberlândia)*; Anderson Santos (Universidade Federal de Uberlândia)

salomao.alves222@gmail.com; santosardr@ufu.br

Resumo: A Bioinformática é uma área interdisciplinar, envolvendo Biologia, Ciência da Computação e Estatística. Possui o objetivo de resolver problemas relacionados à biologia, principalmente na área da biologia molecular, utilizando a informática. A Bioinformática é baseada no armazenamento de grande quantidade de dados biológicos utilizados, por exemplo, em métodos computacionais para localizar padrões. A pesquisa consiste na unificação dos relatórios finais do software Pannotator e do MEDPIPE, que fazem anotação genômica e lista candidatos vacinais, respectivamente. Pois, apesar de existirem programas para realizar a anotação genômica e para descoberta de alvos para vacinas, ainda não há qual faça ambas as funções. A consequência é um tempo maior gasto para descobrir possíveis novas vacinas, bem como lidando com a integração de dados oriundos de fontes diversas. A Anotação Genômica, prediz os genes e provável função para cada proteína derivada do genoma. Para isso, utilizamos o Pannotator, um pipeline automatizado baseado na Web, o qual reduz o trabalho manual e gera relatórios e correções de várias cepas do genoma. Utilizando-se do Pannotator conseguimos realizar uma anotação genômica de forma rápida e mais correta desde que a anotação de referência seja confiável. Já para listagem de bons genes para produzir vacinas e diagnóstico de doenças causadas por agentes infecciosos bacterianos, usamos o MEDPIPE, um servidor para prever proteínas alvo por intermédio da metodologia da vacinologia reversa, é um exemplo de software para elencar candidatos vacinais a partir do genoma de uma bactéria. Ao término da pesquisa, quando você realizar uma anotação genômica via Pannotator, as informações geradas pelo MEDPIPE são integradas com o relatório final do Pannotator. Portanto, o relatório final do Pannotator possuirá possíveis alvos para vacinas, informações sobre localização celular e motivos transmembranares de cada proteína, além das informações tradicionais de provável função e nome do gene provida pelo Pannotator.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Agrupamento de dados via combinação de partições na presença de atributos faltantes

Larissa Pereira (Universidade Federal de Uberlândia)*; Daniel Abdala (Universidade Federal de Uberlândia)

larissa.alves@ufu.br; abdala@ufu.br

Resumo: Os métodos de agrupamento de dados via combinação de partições foram, a priori, idealizados para lidar com conjuntos completos de dados. Infelizmente, é comum encontrar atributos faltantes em conjuntos de dados reais, seja por falha de medição, seja por defeito do sensor. Para lidar com o problema de atributos faltantes, há na literatura diversos métodos de imputação de dados. Apesar de vários métodos serem utilizados para suprir a problemática de valores faltantes, ainda não foi mensurada a qualidade dos resultados dos algoritmos de agrupamento de dados via combinação de partições quando se considera que métodos de imputação foram aplicados na base de dados de entrada. Este trabalho visa avaliar se a utilização de métodos de imputação impactam de maneira negativa o resultado final produzido pelos métodos de agrupamento de dados via combinação de partições. O experimento idealizado visa utilizar um conjunto de quinze bases de dados selecionadas do repositório UCI. Atributos são removidos dos datasets de maneira randômica e progressiva para gerar as lacunas a serem imputadas. Aumenta-se gradativamente a porcentagem de valores faltantes, para que a análise seja feita também sobre o impacto de acordo com a quantidade de lacunas presentes na base de dados. Três métodos de imputação serão aplicados para preencher as lacunas, estimando esses valores. Subsequentemente, os datasets imputados serão processados por cinco algoritmos de clustering que irão compor os resultados de clustering intermediários - o conjunto das funções de particionamento. Finalmente, esse conjunto resultante será processado por três métodos de combinação de partições distintos, que, em sua essência, combinam os resultados de algoritmos de agrupamento, realizando um consenso e escolhendo o resultado mais ajustado. As bases utilizadas serão selecionadas analisando a presença do atributo que representa a classificação real das instâncias. Os datasets com os valores originais também serão processados pelos mesmos algoritmos de clustering e métodos de agrupamento, que atuarão como resultados de controle. A partir destes parâmetros, é possível realizar a análise estatística entre o conjunto de resultados de controle e o conjunto de resultados com datasets imputados para avaliar o impacto geral decorrente da imputação nos algoritmos de agrupamento.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Construção de mecanismo para suportar a predição de tempos de resposta do Cassandra a partir de métricas da infraestrutura openstack

Ian Cunha (Universidade Federal de Uberlândia)*; Rafael Pasquini (Universidade Federal de Uberlândia)

ianresende@ufu.br; rafael.pasquini@ufu.br

Resumo: O Cassandra é um sistema de armazenamento chave-valor, estruturado, distribuído e altamente escalável, projetado para manipular grandes quantidades de dados. Um cliente pode realizar operações de escrita e leitura em um cluster do Cassandra, cada operação realizada possui uma latência, que é o intervalo de tempo entre a requisição e a realização da operação. A carga total em um cluster do Cassandra, que é a quantidade de operações requisitadas pelos clientes em certo momento, varia conforme clientes se conectam ao cluster e realizam operações ou desconectam do cluster e deixam de fazer operações, quando a carga de operações atinge um nível que o cluster não está preparado para atender, espera-se que a latência das operações também aumente. Assim, picos de carga em clusters despreparados podem ocasionar uma perda da qualidade do serviço oferecido. Visando amenizar o impacto ao cliente em decorrência de uma sobrecarga no Cluster Cassandra, faz-se necessário o monitoramento e estudo do comportamento do ambiente frente às variações de carga que ocorrem em situações reais. Para realizar esse estudo foram tomados os seguintes passos: 1- Construção de um ambiente de testes, com um cluster do Cassandra, um cliente Cassandra e um gerador de carga. 2- Construção de uma estrutura para extrair as métricas referentes às operações realizadas pelo cliente e aos recursos de infraestrutura utilizados pelo cluster. 3- Balancear a carga gerada pelo cliente e pelo gerador de carga, para se obter uma configuração de carga que forneça um resultado que evidencie o impacto causado pela variação de carga na performance do cliente. 4- Realizar experimentos que simulam o comportamento de um ambiente DHT real. 5- Com as métricas coletadas gerar um modelo de aprendizado de máquina que possa prever com certa precisão a latência de uma operação conforme as métricas do cluster no dado momento. O ambiente de testes foi criado em máquinas virtualizadas e gerenciadas pelo Openstack, com um nó do Cassandra em cada máquina formando um cluster de 5 nós, uma máquina cliente e uma máquina geradora de carga. O monitoramento e extração de dados no cliente é feito por logs gerados na realização de operações e no cluster pelo software Prometheus. Experimentos com diferentes padrões de carga no cliente e gerador de carga foram realizados para se chegar em um padrão que obtivesse um resultado ótimo na demonstração do comportamento dos componentes do ambiente. Para simular uma carga variável, um padrão de senoide foi configurado no gerador de carga, com clientes sendo criados e removidos durante o experimento. Um padrão de carga alta repentina (flashcrowd) também será configurado. Espera-se que ao aplicar o método de regressão do aprendizado de máquina supervisionado, nos dados coletados, seja possível estimar a latência das operações em vista do estado atual das métricas monitoradas no cluster do Cassandra.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Análise da motivação dos alunos ingressantes em ciências da computação

Felipe Ferreira de Castro (Universidade Federal de Uberlândia)*; João Henrique de Souza Pereira (Universidade Federal de Uberlândia)

felipeaugusto.ferreiradecastro@gmail.com; joaohs@ufu.br

Resumo: Doenças mentais como depressão, estresse, ansiedade, entre outras, tem se tornado crescente na sociedade contemporânea. De acordo com a OMS (Organização mundial de saúde), apud G1, de 2005 a 2015, no Brasil, os casos de ansiedade cresceram 14,9%. Além disso, o país é o primeiro em casos de depressão, com 5,8% da população sendo afetada pela doença. Guimarães (2014) acredita que detectar essas doenças mentais pode levar a um indicativo do estado de saúde de estudantes e jovens profissionais. Assim, esta pesquisa foi proposta com o objetivo de acompanhar o estado emocional dos discentes da Universidade Federal de Uberlândia (UFU), ao longo de seu primeiro e segundo períodos na universidade, utilizando os novos recursos possibilitados pelo BCI (Brain Computer Interface) para coletar dados sobre o estado emocional deles. BCI é uma nova tecnologia que vem sendo desenvolvida à alguns anos e possibilita converter ondas cerebrais em dados computáveis. Desta forma, sendo possível coletar informações sobre as atividades cerebrais do usuário. O aparelho usado nessa pesquisa será o Epoc, desenvolvido pela Emotiv (EMOTIV, 2018), que capta e identifica algumas emoções dos usuários, dentre elas estresse, foco e entusiasmo, as quais serão importantes para esta pesquisa. Visando estimular reações nos participantes, foi feito um questionário sobre a universidade e atividades desenvolvidas no. O discente respondeu equipado com o Epoc, para que fossem captadas suas atividades cerebrais. Desta forma, a resposta oral dada pelo participante foi de importância reduzida, sendo atribuída maior importância às informações coletadas pelo aparelho. Esta pesquisa possui alguns objetivos específicos, os quais são: 1. Coletar dados em diferentes momentos do período; 2. Identificar alterações emocionais nos estudantes ao longo das coletas; 3. Analisar e interpretar os dados coletados; 4. Buscar possíveis causas para as alterações emocionais identificadas, caso essas existam. Até o momento foram realizadas quatro coletas de dados, as quais foram parcialmente analisadas. Os resultados até então obtidos são satisfatórios, tendo sido detectada algumas alterações no estado emocional dos voluntários. Dentre elas, aumento do estresse, redução do relaxamento e do entusiasmo.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Investigando entropia na classificação de tecidos histológicos de mama

Rafael Carvalho (Universidade Federal de Uberlândia)*; Marcelo Zanchetta do Nascimento (Universidade Federal de Uberlândia)

carvarafael@gmail.com; marcelo.zanchetta@gmail.com

Resumo: O câncer de mama é uma das doenças mais comuns em mulheres no Brasil e no mundo. Dentre as possíveis abordagens empregadas no diagnóstico, a análise de imagens histológicas auxiliada por sistemas computacionais de apoio tem se mostrado bastante eficiente para auxiliar os especialistas. Dentre a variedade de métodos de processamento de imagem empregados na análise, as técnicas baseadas em textura têm se tornado referência de bons resultados na classificação dos tumores. Neste trabalho é apresentada uma metodologia de análise de imagens histopatológicas que engloba técnicas de extração de características por meio de histograma de intensidade, aliada com as medidas de entropia de Shannon, Renyi, Tsallis e Kapoor. Finalmente, uma classificação é empregada para avaliação o método baseado no algoritmo Random Forest. Experimentos foram realizados para avaliar a eficiência da metodologia com imagens da base Breakhis, com 7909 imagens histológicas de câncer de mama, coletadas de 82 pacientes, tanto de casos benignos como malignos. Os resultados experimentais mostram uma acurácia de 87,7% e uma área sob a curva ROC de 0,95 com amostragem de validação cruzada com 10-fold. Assim sendo, o sistema proposto permite auxiliar o patologista nos diagnósticos contribuindo para um resultado relevante e confiável.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Estudo de características urbanas por meio da análise de imagens por sensoriamento remoto

Vinicius Sousa (Universidade Federal de Uberlândia)*; André Backes (Universidade Federal de Uberlândia)

viniciusguardieiro@gmail.com; arbackes@yahoo.com.br

Resumo: Com os avanços tecnológicos percebe-se uma melhora na aquisição de imagens de sensoriamento remoto. As imagens de sensoriamento remoto apresentam muitos elementos da superfície terrestre, como rodovias, quarteirões, praças, prédios, áreas verdes, lagos, e outros mais. Por meio dessas imagens é possível obter muitas características de uma região, tornando possível novos estudos através de técnicas de processamento digital de imagem. Esse trabalho teve como intuito processar imagens obtidas pelo Google Earth para a identificação de áreas verde em regiões urbanas e a análise da complexidade de uma área de invasão/ocupação ao longo dos anos. Diferentes metodologias foram estudadas com o intuito de desenvolver um algoritmo capaz de segmentar a região de área verde de forma satisfatória, sendo o melhor resultado obtido através da combinação de limiares e filtros de convolução. Para a análise a invasão/ocupação fez-se necessário o uso da dimensão fractal, pois ela permite qualificar a irregularidade de um objeto e sua homogeneidade.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Aprendizado de máquina na avaliação de risco de crédito financeiro

Janayna Fernandes (Universidade Federal de Uberlândia)*; Murillo G. Carneiro (Universidade Federal de Uberlândia)

fernandesmjjanayna@gmail.com; mgcarneiro@ufu.br

Resumo: Um grande número de instituições tem percebido a importância de suas bases de dados. Essas bases de dados, que geralmente cobrem transações realizadas ao longo de vários anos, podem levar a uma melhor compreensão do perfil de seus clientes e, portanto, vir a corroborar com a qualidade na oferta de novos produtos ou serviços. Nos ambientes de decisão financeira não é diferente já que são utilizadas quantidades substanciais de informações oriundas das mais variadas fontes em um mercado em constante mudança. O aumento do volume de dados armazenado e gerenciado pelas empresas tem se elevado de tal forma que o tratamento dessas grandes bases de dados supera a habilidade humana de entender e de eficientemente lidar com esses dados, levando a uma necessidade crítica de ferramentas e técnicas que sejam capazes de realizar automaticamente análises eficazes e eficientes dos dados para apoiar empresas e indivíduos no planejamento estratégico e na tomada de decisões. O processo geral que possui este fim é geralmente referenciado na literatura como KDD: Descoberta de Conhecimento em Bases de Dados. As razões para as dificuldades da exploração e análise dos dados armazenados são o grande volume dos dados a serem examinados e a natureza dos relacionamentos em si que não são óbvios. Uma ferramenta de descoberta de conhecimento é necessária para auxiliar o tomador de decisão em relação aos pedidos de empréstimo. O processo KDD fornece uma variedade de métodos úteis para descobrir tais relacionamentos em dados históricos, enquanto garante que os relacionamentos descobertos possam ser generalizados para os dados novos (futuros). Ele pode ser descrito em três grandes fases: o pré-processamento dos dados, a mineração de dados e o pós-processamento de dados. A avaliação de risco de crédito é basicamente um problema de classificação que estuda o rendimento e a sua segurança. A decisão de oferecer ou não crédito a um cliente deve ser bastante criteriosa. Dar crédito ao cliente em perspectiva é determinado pelas suas características relacionadas, como idade, renda, solvência ou escolaridade e de características do crédito como tipo de crédito, vencimento, valor do empréstimo e outras características inerentes às operações financeiras. A finalidade dos modelos de avaliação de risco de crédito é classificar os clientes como adimplentes (aceitar) ou como inadimplentes (rejeitar). O conhecimento obtido através do KDD, no final, pode ser usado pelos gerentes de crédito para ajudá-los a rejeitar ou aceitar candidatos. Ao utilizar destas técnicas, haveria menos risco para as empresas financeiras ao prever quais clientes serão bem-sucedidos em seus pagamentos. Consequentemente, mais pessoas poderiam ter acesso a empréstimos de crédito. O objetivo deste trabalho é explorar quatro conjuntos de dados de crédito: alemão, australiano, japonês e taiwanês, obtidos a partir do UCI e comparar a acurácia obtida entre cinco técnicas de mineração de dados: DT, LR, MPL, RF e SVC.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Detecção de comunidades em redes complexas aplicado ao agrupamento de objetos invariantes

Lusmar Freitas (Universidade Federal de Uberlândia - Campus Monte Carmelo)*; Murillo G. Carneiro (Universidade Federal de Uberlândia)

lusmarmf@gmail.com; mgcarneiro@ufu.br

Resumo: O agrupamento é uma tarefa comum em Mineração de Dados e tem por objetivo encontrar grupos de dados caracterizados por algum critério de similaridade. Tais grupos podem trazer informações sobre padrões ocultos nos dados, auxiliando no processo de descoberta do conhecimento. Apesar dos dados serem considerados na forma de vetor de atributos pela maioria dos algoritmos de agrupamento, recentemente o estudo da formação de grupos em redes se tornou um tópico de intensa pesquisa na área, especialmente quando utilizando redes complexas. Redes complexas são uma forma de representação de sistemas reais por meio de grafos, utilizada em diversas áreas do conhecimento como Sociologia, Biologia e Computação. Em redes complexas, comunidade é uma estrutura bastante investigada. Comunidades representam a organização da rede em grupos, os quais podem ser identificados pela existência de muitas arestas conectando vértices de um mesmo grupo e poucas arestas conectando vértices de grupos diferentes. Comunidades são vértices que provavelmente possuem comportamentos ou desempenham funções semelhantes. O processo de determinar as comunidades em uma rede é chamado de Detecção de Comunidades. Como vantagem em comparação com as abordagens tradicionais, os métodos de Detecção de Comunidades examinam também as conexões da rede e não apenas os atributos físicos dos dados (por exemplo, distância ou similaridade). Além disso, estes métodos normalmente formam as comunidades sem estabelecer um número fixo de grupos a priori. A tarefa de agrupamento (ou Detecção de Comunidades) elimina os custos no processo de rotulação das bases de dados, além de facilitar a aplicação em problemas reais. Por isso, ela tem sido utilizada em várias aplicações como exploração de dados científicos, mineração de textos, recuperação de informações, diagnósticos médicos, análises Web, entre outros. Neste trabalho, propõe-se o uso de algoritmos de Detecção de Comunidades para o problema não supervisionado de reconhecimento de padrões invariantes. Dado um conjunto de imagens de objetos fotografados em diferentes posições, ângulos e rotações, o problema consiste em detectar e agrupar imagens relacionadas a um mesmo objeto. Para a realização deste trabalho, foram aplicados algoritmos de Detecção de Comunidades em bases de dados reais disponíveis na literatura. Os métodos de Detecção de Comunidades foram avaliados utilizando um conjunto de métricas de desempenho e comparados aos métodos de agrupamento tradicionais K-means e HDBSCAN. Os resultados obtidos demonstram que os algoritmos de Detecção de Comunidades conseguem atingir boas divisões da rede, se saindo melhor que os métodos tradicionais para a base de dados utilizada. Sendo assim, a utilização de redes complexas se mostra promissora enquanto nova abordagem para o problema de agrupamento de objetos invariantes.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Análise de segurança em sistemas operacionais dedicados para internet of things (IOT)

João de Oliveira (Universidade Federal de Uberlândia)*; Rodrigo Miani (Universidade Federal de Uberlândia)

jpveira@ufu.br; miani@ufu.br

Resumo: A expansão da Internet, a evolução do processamento, a miniaturização da eletrônica foram os principais fatores que motivaram a criação de um paradigma conhecido como Internet das Coisas (IoT -Internet of Things) em que dispositivos de diferentes tamanhos e propriedades como sensores, relógios e câmeras estão diretamente conectados a rede mundial de computadores. Como a complexidade de tais dispositivos estão aumentando e os custos diminuindo, surgiu a necessidade de sistemas operacionais (SOs) que oferecem uma nova camada de abstração para agilizar o desenvolvimento de sistemas mais robustos para IoT. A IoT facilita a troca de informações para suportar serviços da indústria, logística, educação e saúde, facilitando novos modelos de negócios e reduzindo os custos e os riscos, porém, junto com a revolução da tecnologia vem a preocupação com as vulnerabilidades nesses dispositivos. As restrições de recursos em dispositivos dedicados para a IoT fazem com que os protocolos e sistemas de criptografia, que atualmente são considerados robustos e seguros, podem não ser apropriados, ou até inviáveis, para tais dispositivos. A heterogeneidade e a dimensão da rede tornam a segurança em IoT uma questão muito complexa se comparada com a rede tradicional de computadores. Uma forma de tentar evitar algumas falhas é realizando testes no próprio código fonte. Uma opção para apoiar na detecção e correção dessas vulnerabilidades são as ferramentas de análise estática de código. Tais ferramentas testam o código sem executá-lo, analisando o código fonte ou tendo por base uma entrada especificada, automatizando a detecção de violações de regras em um projeto, forçando o uso de boas práticas de programação. Diversas pesquisas utilizam dados coletados por essas ferramentas para auxiliar a compreensão da qualidade do código. O objetivo do presente trabalho é analisar as funções inseguras presentes nos códigos fontes dos seguintes sistemas operacionais para IoT: Contiki-OS, FreeRTOS, Mbed-OS, RIOT, Tiny-OS, TizenRT e Zephyr. Tais sistemas são populares e possuem o código aberto, o que facilitou a coleta dos dados. A ferramenta utilizada para a análise estática do código-fonte foi o flawfinder, que possui foco em funções inseguras. Dentre as análises realizadas, pode-se destacar a avaliação da frequência de ocorrência das funções inseguras, o número de vulnerabilidades associadas a tais funções inseguras, a porcentagem de funções inseguras por linha de código-fonte e identificação de funções inseguras com base no CWE (Common Weakness Enumeration).

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Gerontecnologia: um estudo de caso para inclusão digital

Giullia Menezes (Universidade Federal de Uberlândia)*; Renan Cattelan (Universidade Federal de Uberlândia)

giulliam@gmail.com; renan@ufu.br

Resumo: A exclusão digital é um problema que aflige parte significativa da população, com incidência maior entre o público idoso, justamente um dos segmentos que mais cresce considerando-se o aumento da expectativa de vida. Os idosos precisam de meios adequados para manter suas habilidades funcionais, fazendo com que ações de inclusão digital específicas para o público idoso sejam cada vez mais necessárias e urgentes, de modo a melhorar aspectos de integração social e de qualidade de vida. Com o objetivo de auxiliar o público idoso a se adaptar às novas tecnologias e poder delas usufruir, o CompPET desenvolve, já há 7 anos, diversas iniciativas para promover a inclusão digital de idosos. A metodologia adotada consiste na oferta de cursos e oficinas voltadas ao público idoso, com as seguintes características: flexibilidade parcial de conteúdo; material didático próprio, gratuito e livremente disponível, complementado por conteúdo desenvolvido por especialistas do NIC.br especialmente para esse público-alvo; aulas dinâmicas, com colaboração em pares; aproveitamento de infraestrutura dos laboratórios de informática da universidade; emprego de recursos de acessibilidade; e formação de parcerias. Um novo desafio, proposto neste projeto, tem sido o de renovar essas metodologias, de forma com que o aprendizado se torne mais interessante e de fácil entendimento, para que o conteúdo repassado seja realmente absorvido e que os idosos se capacitem para utilizar a tecnologia em seu cotidiano.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

V2C – protótipo de veículo conectado

Guilherme Henrique de Araújo Santos (Universidade Federal de Uberlândia)*; Luiz Theodoro (Universidade Federal de Uberlândia)

guilhermeher10@gmail.com; luiz.theodoro@ufu.br

Resumo: V2C (Vehicle-to-Cloud), em português Veículo para Nuvem ou protótipo de veículo conectado, se trata de uma interação entre um veículo diretamente com a nuvem, possibilitando que motoristas e passageiros acessem aplicativos de informação, navegação e entretenimento diretamente de um display instalado no carro ou até mesmo de um aparelho remoto, como os atuais smartphones. O protótipo de veículo conectado em si, seria uma forma de também fazer com que qualquer tipo de problemas nos carros sejam identificados com maior assertividade e eficiência, poupando tempo e mão de obra. O processo da nuvem entra para garantir uma opção mais segura entre fornecedor e cliente, com a vantagem de acessos em qualquer lugar do planeta. O projeto envolve cerca de 4 áreas da engenharia, Engenharia da Computação (Engenharia de software), Engenharia Eletrônica e de Telecomunicações, Engenharia de Automação e Engenharia Mecânica. O objetivo geral desse projeto de pesquisa é aprofundar na proposta de NaaS - Network as a Service, envolvendo IoT - Internet das Coisas vinculada aos benefícios do Cloud Computing por meio do desenvolvimento de uma solução que integre ambas as tecnologias. Dessa maneira, aproveitar as possibilidades de provisionamento usando estas novas tecnologias para gerar benefícios como melhor desempenho, vazão de dados escalável, menor latência e maior simplicidade na configuração. Como entrega final, a proposta é desenvolver, utilizando um veículo com pouquíssima inovação tecnológica, um computador de bordo acoplável a um veículo com inúmeros serviços amparados na coleta de dados por sensores como medição de pressão, de volume de combustível, de velocidade ou controles como ré, presença e outros com hospedagem em ambientes virtualizados em nuvem que processaria a inteligência do sistema. A ideia é que qualquer equipamento possa ser monitorado por meio de sensores conectados à uma plataforma que permita funcionalidades comuns aos veículos atuais. A aplicação poderá ser para Smartphone que atuará sobre uma camada de software interagindo com um sistema de sensores e dispositivos especiais, utilizando sistemas de comunicação mais atuais como o LoRa (LONg RANGE) para maior eficiência e autonomia afim de melhorar a experiência da viagem.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Estudo comparativo de técnicas do estado da arte para predição de séries temporais não-estacionárias.

Higor Emanuel Silva (Universidade Federal de Uberlândia)*; Marcelo Albertini (Universidade Federal de Uberlândia)

higoress@gmail.com; albertini@ufu.br

Resumo: Séries temporais são conjuntos de observações de um fenômeno ordenados no tempo. Devido a essa ordem os dados possuem dependência, onde os dados mais próximos afetam mais o valor atual do que os outros. Uma das aplicações mais importantes do estudo de séries temporais é a predição. Predizer a probabilidade de chuva no próximo dia baseado nas condições climáticas da última semana, ou o preço de uma ação na bolsa de valores de acordo com as variações na última hora, ou até mesmo estimar os horários de pico em um hospital para alocação de mais enfermeiros para o pronto atendimento. Com esse intuito, para realizar esta tarefa é preciso criar modelos que tentam explicar de maneira mais precisa possível como esse fenômeno comportar-se-á. As técnicas bem estabelecidas hoje em dia funcionam muito bem para séries temporais estacionárias, onde os dados permanecem em equilíbrio estatístico, de média ou variância, através do tempo. Para as séries não-estacionárias, onde os dados não possuem esse equilíbrio estatístico em pelo menos um dos intervalos de tempo, o trabalho se torna um pouco mais difícil. As soluções disponíveis hoje em dia, são principalmente ARIMA, Prophet e técnicas de Machine Learning. O modelo ARIMA é subdividido em três partes: AR(p) de auto-regressivo, onde é feita uma regressão linear com os p últimos termos, MA(q) de médias móveis, que faz a regressão linear com a média dos q últimos erros e I(d) que elimina a não-estacionariedade de algumas séries aplicando d diferenciações sucessivas. Entretanto, esse modelo faz suposições muito fortes sobre a distribuição dos erros, como sendo pertencentes a uma distribuição Gaussiana, o que não é o caso para a maioria das séries do mundo real. Sobretudo, modelos lineares não criavam modelos precisos para dados complexos. E para isso os modelos de Machine Learning eram mais adequados. Devido a natureza dos dados, com dependência temporal, o uso das redes neurais Long Short Term Memory (LSTM) gerou bons resultados. Contudo, o processo de aprendizado supõe a estacionariedade da série e que os dados sejam independentes e identicamente distribuídos (I.I.D.). Com a finalidade de melhorar tais resultados, Kuznetsov e Mohri (2018) propuseram um novo método onde suposições mais brandas são feitas sobre a origem dos dados e diferentemente do modelo ARIMA, que usa o método de máximo verossimilhança, esse método resulta em um problema de otimização convexa onde o máximo global pode ser eficientemente encontrado. Deste modo, nós implementaremos um algoritmo baseado em Kuznetsov e Mohri (2018), e um ambiente de testes para comparar a performance do nosso algoritmo com estes tido como estado da arte: ARIMA, Prophet e Machine Learning com LSTM. Os critérios de análise serão: complexidade computacional, erro quadrático médio e erro médio absoluto.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Estruturas de indexação baseadas em mapeamentos unidimensionais para otimização de consultas por similaridade

Tércio de Melo Alves Júnior (Universidade Federal de Uberlândia)*; Humberto Razente (Universidade Federal de Uberlândia)

humberto.razente@gmail.com; terciouz.jr@hotmail.com

Resumo: Em geral, as aplicações tradicionais que manipulam dados numéricos e textuais realizam consultas baseadas em igualdade e ordem total em sistemas de gerenciamento de bancos de dados otimizados por meio de árvores B+ e hash. Entretanto, para dados de natureza multimídia, consultas baseadas em distância são de maior interesse. Uma consulta por similaridade é geralmente definida por um elemento de consulta e uma restrição baseada na distância em relação ao elemento de consulta. Para a realização eficiente de consultas por similaridade é necessário armazenar os dados em métodos de acesso apropriados para a realização de consultas baseadas em distâncias. Este projeto de iniciação científica aborda o desenvolvimento de um método de acesso métrico baseado em árvores B+ por meio de mapeamentos unidimensionais obtidos a partir de conjuntos estáticos de pivôs para melhorar o desempenho de consultas por similaridade, em especial, consultas aos k-vizinhos mais próximos de modo exato, nas quais a desigualdade triangular é empregada para eliminar porções de dados que certamente não fazem parte da resposta. A validação experimental é realizada por meio da comparação com os métodos disponíveis na literatura, entre eles o iDistance e Omni-B-Forest, sendo que os resultados iniciais são promissores.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Análise do tráfego de rede gerado por veículos aéreos não tripulados comerciais

Leandro Carneiro (Universidade Federal de Uberlândia)*; Rodrigo Miani (Universidade Federal de Uberlândia)

leandro-ebc@hotmail.com; miani@ufu.br

Resumo: Os veículos aéreos não tripulados (VANTs) também conhecidos por drones se tornaram populares e bastante utilizados em atividades de vigilância, transporte de cargas, captura de vídeos ou fotos, agricultura, apagar incêndios e também fins militares o que gera um grande benefício por simplificar tarefas para os usuários e até impedir que fiquem expostos a situações de risco. Porém, o uso inadequado do equipamento pode ocasionar diversos problemas como: invasão de privacidade, acidentes e catástrofes onde drones foram modificados para portar e usar armas de fogo. Os drones armazenam informações sensíveis como imagens, vídeos, localização e áudios que são coletados e armazenados durante sua utilização e a presença de vulnerabilidades e brechas nesses sistemas, possibilitam que atacantes explorem e roubem os dados do aparelho. Uma outra preocupação está relacionada com o comportamento das próprias empresas que comercializam os drones. Um caso recente aconteceu com a empresa DJI. Ela foi acusada por militares norte-americanos de armazenar dados dos usuários em servidores localizados na China, sem o consentimento dos mesmos. Dada a importância das informações para todos os envolvidos e a questão da privacidade dos usuários, realizar estudos para identificar como os drones gerenciam esse tipo de situação se torna uma tarefa relevante, não somente para a sociedade, mas também para a comunidade científica. O objetivo deste trabalho é conduzir uma análise do tráfego de rede enquanto o usuário pilota dois modelos comerciais de drones produzidas pela empresa DJI: Phantom 4 e Mavic 2. O ambiente experimental possui as seguintes características: os drones são controlados por um controle remoto e um smartphone conectados via USB e há um aplicativo desenvolvido pela DJI onde é possível utilizar algumas funcionalidades que permite gerenciar o VANT. O processo de interceptar os pacotes que trafegam na rede é feito com o auxílio de um analisador de pacotes como o Wireshark. A partir da análise dos arquivos gerados pelos Wireshark é possível verificar o tipo de tráfego de rede gerado pelo drone e identificar pontos como: destino dos pacotes, tamanho médio dos pacotes, tipos de protocolos de rede, presença ou ausência de dados cifrados, entre outros. O trabalho também irá confrontar as informações obtidas com os termos de uso propostos para cada tipo de drone analisado.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Progressive web apps: características e aplicações

Fabrizio Ziliotti (Universidade Federal de Uberlândia)*; Renan Cattelan (Universidade Federal de Uberlândia)

ffz.ziliotti@gmail.com; renan@ufu.br

Resumo: Uma Progressive Web App (PWA ou, em Português, Aplicativo Web Progressivo) engloba diversas tecnologias para criação de um website capaz de se comportar como um aplicativo nativo de um dispositivo móvel. Esse novo comportamento foi possibilitado graças às funcionalidades que os navegadores atuais estão oferecendo, como: acesso a aplicação de forma offline, recursos de geolocalização, notificações push, além da possibilidade de instalação do aplicativo na área de trabalho. O objetivo desta pesquisa é analisar e implementar aplicações que contemplem as características de uma PWA, verificando e aprimorando os scripts de service workers, que são a base do funcionamento inteligente da PWA. Ademais, estão sendo investigadas as melhores práticas utilizadas pelas principais bibliotecas e frameworks Web, como: ReactJS, Angular, VueJS, Polymer e Gatsby. Essas práticas reúnem a implementação de sites e sistemas Web com alto desempenho, cacheamento inteligente para os recursos utilizados pelo aplicativo, implementação de uma metodologia para funcionamento offline e entrega de boa experiência ao usuário. Dessa forma, podem ser levantadas informações a respeito das diferenças e semelhanças entre aplicativos nativos e PWAs. Além disso, será desenvolvido o blog Printf, como prova de conceito, bem como serão feitas análises e implementações de estruturas de aplicações Web utilizando bibliotecas JavaScript voltadas para a construção de aplicativos e sistemas Web. Para estudo e auditoria das aplicações está sendo utilizado o software Lighthouse, ferramenta automatizada de código aberto que aprimora a qualidade de apps da Web. Ele pode ser executado como extensão do Chrome ou na linha de comando com Nodejs. Com a pesquisa, ainda em andamento, foi possível observar que a comunidade que apoia as grandes bibliotecas e frameworks populares, como React, Angular, Vue, responsáveis pela construção de sistemas Web, está se adaptando às mudanças e disponibilizando o suporte a PWAs em seus códigos, contribuindo para a propagação e avanço da tecnologia em grande variedade de segmentos de desenvolvimento.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Uso de modelos probabilísticos para melhorar a eficiência da exploração de sequências de otimização do compilador

Domitila Crispim (Universidade Federal de Uberlândia)*; Luiz Martins (Universidade Federal de Uberlândia)

domitilacrispim@gmail.com; lgamartins@ufu.br

Resumo: O desempenho do código é um fator importante no projeto de software. Em sistemas embarcados, que apresentam limitações arquiteturais como restrições de memória e processamento, esse desempenho está diretamente associada à adequação do código aos recursos disponíveis. Durante o processo de compilação, no qual ocorre a transformação do código descrito pelo programador naquele efetivamente executado na máquina, passos de otimização podem ser aplicados, propiciando oportunidades de aperfeiçoamento do código final. Compiladores comerciais costumam adotar níveis de otimização, que são sequências fixas escolhidas com base na sua capacidade de otimizar um grande variedade de códigos (representatividade) e na experiência dos desenvolvedores do compilador. Devido a sua característica genérica, elas não garantem que o código seja realmente otimizado, nem que o melhor desempenho será alcançado. Encontrar sequências específicas para o código a ser compilado é uma tarefa importante, mas muito complexa. O espaço de busca é enorme, dado que existem centenas de passos de otimização e que a ordem de aplicação também afeta o desempenho final do código. Nesse cenário, a busca exaustiva é inviável, sendo necessário o uso de heurísticas de busca para encontrar boas sequências de otimização de acordo com o código compilado e a arquitetura alvo. Após experimentos preliminares no ambiente de compilação LLVM, observou-se que a geração aleatória, adotada em muitas heurísticas, compromete a eficácia da exploração de sequências, dado que ocasionam um número elevado de soluções inválidas, ou seja, sequências que não podem ser aplicadas ou que geram códigos executáveis incorretos. Para lidar com esse problema, este trabalho propõe o uso de um modelo probabilístico para direcionar a geração de sequências de otimização válidas. O modelo probabilístico adotado é gerado a partir de dados históricos de compilação, ou seja, das sequências encontradas para os códigos já compilados. Ele pode ser representado como um grafo direcionado, no qual cada vértice representa um passo de otimização e as arestas representam a frequência histórica da ordem de aplicação. A geração de uma sequência consiste em escolher aleatoriamente o tamanho da sequência e o primeiro passo de otimização a ser aplicado, bem como utilizar o modelo construído para escolher os demais passos de acordo com as probabilidades associadas a cada aresta que se origina no passo atual. Essa abordagem visa eliminar ou, pelo menos, diminuir significativamente a quantidade de soluções inválidas avaliadas durante a exploração e, conseqüentemente, melhorar a eficiência e eficácia do método de exploração utilizado. Resultados experimentais mostraram que o uso da nossa abordagem eliminou a ocorrência de sequências inválidas e resultou em um speedup médio de 1,45x sobre a exploração baseada no mesmo método sem a nossa abordagem, comprovando a efetividade do método proposto.

Trilha: Trabalhos de Graduação

*Apresentador(a) do trabalho.

Parte II

TRABALHOS DE MESTRADO

Avaliação da capacidade de retenção de água em amostras de carne suína

Renan de Andrade (Universidade Federal de Uberlândia)*; André Backes (Universidade Federal de Uberlândia)

andraderenan@yahoo.com.br; arbackes@yahoo.com.br

Resumo: A carne suína é uma das fontes de proteína animal mais consumidas no Brasil e no mundo. Diante disso, é extremamente importante que a carne oferecida aos consumidores passe por um rigoroso processo de controle de qualidade. Normalmente, a qualidade da carne suína está associada a medidas de características desejadas e valorizadas pelo consumidor. Dentre as várias características analisadas para se obter um produto de alta qualidade, a capacidade de retenção de água (CRA) é uma das mais utilizadas. Os níveis de concentração de água na carne suína estão relacionados diretamente com o seu aspecto visual, maciez, suculência, firmeza e rendimento durante o preparo. A técnica CRA estima a quantidade de água de uma amostra ao medir o volume de suco liberado após a compressão da amostra. No entanto, esse processo é realizado manualmente por especialistas, o que faz dela uma operação demorada, imprecisa, variável, cara e vulnerável a falhas e ao cansaço humano. Tendo em vista a necessidade de automatização desse processo e a escassez de pesquisas que utilizam técnicas de processamento digital de imagens para segmentar e quantificar CRA, foi desenvolvido um método de segmentação automática capaz de quantificar as CRAs de amostras de carnes suínas. A eficiência do método proposto é avaliada a partir da realização de experimentos onde a segmentação das amostras e a precisão das CRAs extraídas são analisadas. Em ambos os experimentos, o método proposto obteve resultados satisfatórios. Sendo assim, este trabalho poderá auxiliar os profissionais da área de Ciências Veterinárias a realizar uma quantificação de CRA mais eficiente, prática e com menor custo.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Algoritmos evolutivos aplicados ao problema de escalonamento de tarefas em sistemas multiprocessados

Eduardo Silva (Universidade Federal de Uberlândia)*; Paulo H. R. Gabriel (Universidade Federal de Uberlândia)

eduardo.iftm@gmail.com; paulohrgabriel@gmail.com

Resumo: Ao dividirem o uso de um processador, diversas tarefas em um sistema computacional devem obedecer uma ordem de execução estabelecida, de forma a proporcionar a sua completa execução. Essa sequência de execução tem sua complexidade aumentada quando as tarefas encontram um cenário com múltiplos processadores, seja por unidades processadoras multi-núcleos ou por arranjos de computadores interligados em rede. O escalonamento de tarefas em sistemas multiprocessados (ETSM) é um problema da classe NP-difícil que consiste em atribuir tarefas sobre recursos computacionais, de maneira a otimizar uma ou múltiplas métricas de desempenho. Restrições de precedência, heterogeneidade dos recursos computacionais e comunicações entre as tarefas são alguns exemplos de características que aumentam a complexidade do problema. Dentre variadas propostas de solução, destaca-se o uso dos algoritmos evolutivos (AE). Inspirando-se no paradigma evolucionista, os AEs definem codificações que padronizam as variáveis de decisão utilizadas em cada possível solução. Ao medir a qualidade de cada solução por meio de uma função objetivo, AEs podem simular um processo de evolução natural ao criar coleções de soluções que podem ser modificadas e evoluídas paralelamente. Os operadores de recombinação e mutação são essenciais na busca de melhores soluções ao promover combinações entre soluções ou modificações aleatórias. Além de restringir o espaço de busca para soluções do problema, a codificação utilizada pelo AE também determina a complexidade dos operadores utilizados. Este trabalho apresenta uma Revisão Sistemática da Literatura (RSL), processo estruturado que propõe planejamento, busca, avaliação e sumarização de trabalhos científicos associados a uma determinada questão de pesquisa. As buscas são baseadas em palavras-chave empregadas em bases acadêmicas. Em seguida, seleciona-se trabalhos por um processo de avaliação. A questão principal adotada é: “Quais as principais abordagens evolutivas para resolver o problema do ETSM?”. Existem, também, questões secundárias como: “Quais os operadores, representações e função objetivo utilizados?”. Inicialmente, as buscas apresentaram 368 trabalhos relacionados. Destes, 120 trabalhos foram pré-selecionados pela leitura do resumo. Posteriormente, uma leitura das seções, voltadas à especificação do problema e ao detalhamento das técnicas, definiu uma coleção inicial específica com 36 trabalhos. A partir desses artigos, outra coleção de 29 trabalhos foi relacionada com base nas referências bibliográficas utilizadas. Ao todo, 65 trabalhos foram analisados sob às questões propostas na RSL. Variadas codificações foram levantadas, bem como seus respectivos operadores. Elencou-se, também, diferentes métricas de desempenho. Ao final, as codificações mais utilizadas terão suas representatividades comparadas, de forma que o resultado dessa avaliação poderá ser utilizado na definição de novas propostas evolutivas.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Aprimorando o desempenho na cobrança de faltas de agentes jogadores de FIFA por meio de técnicas de detecção de objetos

Matheus Faria (Universidade Federal de Uberlândia)*; Rita Julia (Universidade Federal de Uberlândia); Lídia Tomaz (Instituto Federal do Triângulo Mineiro)

matheus_prandini@ufu.br; ritasilvajulia@gmail.com; ldbononi@gmail.com

Resumo: Como os agentes inteligentes têm sido cada vez mais necessários para resolver problemas do dia a dia, o Aprendizado de Máquina (AM), com ênfase especial nos algoritmos de Aprendizagem por Reforço Profunda (ARP), tem sido um excepcional campo científico de pesquisa. Diversos problemas abordados por esses agentes, tais como jogos digitais e sistemas autônomos, exigem a capacidade de Visão Computacional (VC) como um recurso adicional a ser acoplado ao seu processo de aprendizagem. As Técnicas de Detecção de Objetos (TDOs) representam uma importante ferramenta auxiliar da VC para melhorar a habilidade dos agentes de perceber o ambiente em que atuam, o que aprimora suas capacidades de tomada de decisão. Os jogos digitais destacam-se como um estudo de caso nas pesquisas baseadas em AM+VC pelas seguintes razões: além de apresentarem dificuldades técnicas desafiadoras, possuem grande impacto no cenário econômico atual. Embora várias dessas pesquisas usem algoritmos de ARP como abordagem de aprendizado, poucas delas usam as TDOs para melhorar sua estratégia de representação de estado. Motivado por esses argumentos, o presente trabalho propõe dois agentes baseados em ARP+TDO para lidar com a tarefa de cobrar faltas no ambiente do jogo digital FIFA. As principais contribuições aqui são: testar o desempenho da versão aprimorada da TDO nomeada como "MobileNetV2" em ambientes dinâmicos; e melhorar o desempenho de um agente na tarefa aqui abordada por meio da inclusão da barreira como um novo elemento a ser detectado pela TDO "MobileNetV2". Os resultados confirmam a eficiência do "MobileNetV2" em ambiente dinâmico, bem como a melhora no desempenho dos agentes com o uso dessa técnica.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Desenvolvimento de algoritmos e técnicas em Bioinformática para estudo de genomas de patógenos humanos procariotos

William Ferreira (Universidade Federal de Uberlândia)*; Anderson Santos (Universidade Federal de Uberlândia)

williamferreira_7@hotmail.com; santosardr@gmail.com

Resumo: Na pesquisa por drogas e vacinas contra doenças causadas por organismos bacterianos utiliza-se frequentemente redes PPIs (do inglês Protein-Protein Interaction) como ferramenta de análise para identificação de proteínas com potencial de utilização na confecção de fármacos. No entanto, atualmente não há muitas alternativas de software para realizar a predição de novo de PPIs para um organismo cuja sequência genômica acabou de ser elucidada. Este trabalho propõe o desenvolvimento de um programa de computador para predição de redes PPIs pelo método ab initio, ou seja, sem uso de homologia entre proteínas.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Predição de porções de proteínas que estimulam o sistema imunológico adaptativo utilizando uma rede neural convolucional (Deep Learning) programada em Lisp

Rafael Queiroz (Universidade Federal de Uberlândia)*; Anderson Santos (Universidade Federal de Uberlândia)

rafael.queiroz@ufu.br; santosardr@gmail.com

Resumo: Esta pesquisa tem o propósito desenvolver uma ferramenta capaz de identificar epítomos que geram resposta imunes. Esta identificação é importante para geração de vacinas, diagnósticos e soroterapias. Para isso realizaremos o refactoring do NetMHC que utiliza técnicas de aprendizado de máquinas para predizer ligações peptídica do Complexo principal de histocompatibilidade (MHC). Esta pesquisa também possui o propósito de contribuir com o projeto Mature Epitope Density (MED).

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Classificação multirrótulo de alto nível por meio do uso de redes complexas

Vinicius Resende (Universidade Federal de Uberlândia)*; Murillo G. Carneiro (Universidade Federal de Uberlândia)

resendeviniciush@gmail.com; mgcarneiro@ufu.br

Resumo: Em aprendizado de máquina, um classificador é um modelo treinado capaz de associar de forma automática um conjunto de valores de entrada as suas respectivas saídas (também denominadas como classe ou rótulo). A versão mais conhecida do problema de classificação (comumente chamada na literatura de classificação monorrótulo), assume que as entradas devem ser associadas a um único e exclusivo rótulo, tornando o classificador incapaz de resolver problemas onde suas instâncias possam estar associadas a múltiplas saídas simultaneamente (e.g., diagnóstico de doenças, categorização de textos, classificação de músicas, etc). Para contornar tal limitação, algoritmos de classificação multirrótulo estão sendo desenvolvidos e estudados desde a última década, onde a maior parte, trata-se de adaptações de versões monorrótulo. Em comum, as técnicas de classificação mono e multirrótulo classificam tomando como base apenas aspectos físicos dos dados (e.g., distância, similaridade ou distribuição), não levando em consideração características topológicas, como por exemplo, padrão de formação. Inspirado em recentes avanços no uso de redes complexas para classificação monorrótulo, este trabalho visa investigar uma solução capaz de analisar tanto as características físicas, quanto a estrutura topológica dos dados, combinando classificadores multirrótulo existentes com um classificador de alto nível construído com o uso de redes complexas. Experimentos realizados em bases de dados artificiais e reais mostram que tal abordagem é promissora.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Descoberta antecipada de exploits usando redes sociais

Daniel Sousa (Universidade Federal de Uberlândia)*; Rodrigo Miani (Universidade Federal de Uberlândia); Elaine Faria (Universidade Federal de Uberlândia)

danielnovato@gmail.com; miani@ufu.br; elaine@ufu.br

Resumo: No contexto da segurança de computadores, vulnerabilidades são falhas de softwares que permitem a agentes maliciosos violarem políticas de segurança. De maneira geral, uma vulnerabilidade possui ciclo de vida onde, a partir de sua descoberta, inicia-se uma condição de corrida entre desenvolvedores/usuários e atacantes. Por um lado atacantes desejam construir ferramentas ou métodos (chamados de exploits) e realizar ataques antes que os usuários tenham a chance de instalarem correções, por outro, usuários precisam que patches de correção sejam criados pelos desenvolvedores do software afetado para que possam avaliar os impactos de uma atualização e, finalmente, atualizarem seus sistemas e não mais estarem vulneráveis. É grande o volume de vulnerabilidades descobertas e divulgadas todos os anos, somente em 2018 foram 16.555, de acordo com a National Vulnerability Database (NVD), volume que dificulta o trabalho de desenvolvedores de software e gerenciadores de sistemas. Apesar disso, nem toda vulnerabilidade é explorada, estudos mais recentes mostram que muitos exploits são utilizados apenas como prova de conceito mas não oferecem risco provável ao usuário, por serem de uso impraticável. Levando isso em conta, um estudo de 2015 mostrou haver evidências de que apenas 1,3% das vulnerabilidades catalogadas no ano anterior foram exploradas no mundo real. Métodos convencionais de se mensurar a probabilidade de uma vulnerabilidade ser explorada utilizam informações sobre a própria natureza da falha, através de métricas de severidade como o Common Vulnerability Scoring System (CVSS). Porém, redes sociais como o Twitter e Facebook têm se tornado ambientes de discussão de temas muito diversos, incluindo assuntos relacionados a segura de computadores. Desenvolvedores, analistas de segurança e mesmo hackers compartilham informações que podem ser úteis a área. O objetivo geral deste trabalho é propor e aplicar um método que supere a precisão, revocação e o tempo de resposta dos classificadores já existentes para detecção de vulnerabilidades exploradas através do uso de bases de dados mais amplas do que aquelas usadas em trabalhos anteriores. Deverão ser consideradas apenas as vulnerabilidades já divulgadas e catalogadas (com CVE). Para verificar a eficácia do método, deverá ser construída uma solução que classifique vulnerabilidades como "exploradas" ou "não exploradas" quanto ao seu uso por agentes maliciosos no mundo real.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Aprendizado de ranking de entidades aplicado aos dados do governo brasileiro

Paulo Henrique Soares (Universidade Federal de Uberlândia)*; Marcelo Albertini (Universidade Federal de Uberlândia)

phmaias@gmail.com; albertini@ufu.br

Resumo: Com o crescimento da quantidade de informações referente a transparência governamental disponíveis nos últimos anos devido às exigências legislativas, o acesso a informação desejada torna-se cada vez mais difícil. Buscadores tradicionais como Google, Yahoo e Bing retornam as informações desejadas ordenada pela relevância do documento perante a consulta informada. A área cujo objetivo é retornar os documentos relevantes ao usuário é conhecida como Recuperação de Informação a qual pode ser auxiliada por algoritmos de aprendizado de máquina para melhorar a ordenação dos documentos, denominada nesse contexto como Aprendizado de Ranking. Existem na literatura diversos algoritmos para resolver problemas de Aprendizado de Ranking o qual cada um busca resolver o problema de ordenação da melhor forma possível. No contexto de documentos governamentais observa-se a possibilidade de identificar quais são as principais entidades presentes nos documentos a fim de agregá-las na pontuação de ranking. Esse trabalho visa verificar uma maneira de obter uma ordenação dos documentos disponíveis no Portal Dados do Governo Brasileiro utilizando Aprendizado de Ranking em conjunto com as informações de entidades destas bases, que são neste contexto dados não-estruturadas, semi-estruturadas e tabulares. Para atingir tal objetivo utilizaremos técnicas disponíveis no estado da arte para reconhecimento de entidade nomeadas e modelos de otimização convexa para modelar o processo de aprendizado de ranking.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Orquestração osmótica de cloud slices

Aryadne Guardieiro Pereira Rezende (Universidade Federal de Uberlândia)*; Rafael Pasquini (Universidade Federal de Uberlândia); Raquel Fialho Lafetá (Universidade Federal de Uberlândia)

aryadne.guardieiro@ufu.br; rafael.pasquini@ufu.br; raquel.rafielho@gmail.com

Resumo: Criando um novo paradigma de slice como serviço, o projeto Novel Enablers for Cloud Slicing (NECOS) visa prover recursos de uma federação de provedores de infraestrutura de maneira uniforme, integrada e softwarizada. Uma slice, ou fatia, pode ser vista como um conjunto de recursos físicos ou virtuais (rede, processamento e armazenamento) que podem acomodar componentes de um serviço, de maneira integrada e independente de outras slices. Na proposta do projeto, primeiro, o cliente fornece uma descrição da slice ao sistema NECOS. Depois de alocado esse conjunto de recursos, de um ou mais provedores, o cliente poderá prover seus serviços aos seus usuários finais. Devido a natureza dinâmica da maioria das aplicações hospedadas na nuvem, a demanda por recursos poderá sofrer variações para mais ou para menos. Para manter o Service Level Agreement (SLA) contratado e os custos do cliente proporcionais à demanda, é necessário um orquestrador inteligente que gerencie a alocação e desalocação de recursos de maneira dinâmica. Para tal, esta dissertação explora o conceito químico da osmose na recomodação de recursos dentro da slice, migrando recursos excedentes de uma parte pouco utilizada para uma parte que esteja sobrecarregada, equilibrando recursos e demanda entre as partes da slice. Em um ciclo de controle infinito, tal mecanismo seguirá os passos: 1 - Estimativa do SLA que o usuário final está recebendo, utilizando métricas da infraestrutura e algoritmos de aprendizado de máquina do tipo regressão. 2 - Observação do histórico de SLAs estimados durante uma janela temporal, a fim de detectar se há uma tendência de degradação da qualidade do serviço, utilizando algoritmos de reconhecimento de padrões. 3 - Análise da configuração atual da slice, para rebalancear os recursos entre suas partes, migrando recursos excedentes de uma parte pouco utilizada para uma parte que necessite mais daquele recurso, com a ajuda de algoritmos genéticos. 4 - Por último, execução do plano criado no passo 3, disparando as ações sugeridas para os receptores do NECOS dentro dos provedores, que executarão as modificações das configurações dos recursos em si. Para testar a abordagem osmótica de orquestração, foi criada uma plataforma de testes na qual um serviço de vídeo sob demanda com qualidade adaptativa (usada como SLA) é disponibilizado. Para disponibilizar o serviço são utilizadas tecnologias como: Protocolo MPEG-Dash, Docker, Kubernetes e Openstack, enquanto que para a monitoria da infraestrutura é empregado o software Prometheus. No momento, está sendo desenvolvido o passo 1 do ciclo anterior. Através da aplicação de cargas sintéticas variáveis, coleta de dados da infraestrutura e de um usuário final, algoritmos de regressão permitem gerar um modelo que represente as oscilações da qualidade do serviço durante as oscilações da demanda. Em seguida, algoritmos de regressão para estimativa do SLA serão analisados, completando a fase 1 do ciclo.

Trilha: Trabalhos de Mestrado

*Apresentador(a) do trabalho.

Parte III

TRABALHOS DE DOUTORADO

CryptoCeA: criptografia com autômatos celulares

Everton Lira (Universidade Federal de Uberlândia)*; Gina Oliveira (Universidade Federal de Uberlândia)

evertonlira@gmail.com; gina@ufu.br

Resumo: No contexto atual onde a segurança digital é altamente relevante para serviços bancários, troca de dados confidenciais e acessos restritos a senha, a criptografia é uma ciência indispensável como ferramenta de provisão de tal segurança. A evolução do poder computacional exige uma evolução contínua dos métodos criptográficos e, nesse sentido, uma abordagem viável para prover tais métodos é pelo uso dos autômatos celulares, que são sistemas discretos descritos inicialmente por John von Neumann em 1966. Tais sistemas fazem uso de combinações entre regras e indivíduos relativamente simples para expressar comportamentos que podem ser vistos como complexos. O modelo de criptografia simétrica HCA, concebido por Heverton Barros e Gina Oliveira, foi descrito em 2007 na dissertação de Heverton. Esse modelo considerava o texto claro criptográfico (em forma binária) como sendo o reticulado inicial de um autômato celular e, tomando regras de evolução como equivalentes à chave criptográfica, correspondia a evolução do autômato a um procedimento criptográfico resultando em texto cifrado. Este mesmo modelo foi a base para concepção do CryptoCeA (nome temporário), que dados os bons resultados do HCA, faz uso de inovações sob a premissa de aumentar a eficiência computacional e segurança criptográfica do método original. Resultados iniciais foram considerados positivos, porém resultados mais apurados estão sendo extraídos do novo método no estágio atual do estudo.

Trilha: Trabalhos de Doutorado

*Apresentador(a) do trabalho.

Open-and-secure data framework (OSDFramework): a pragmatic proposal for building new applications from existing iot applications

Rodrigo Francisco (Universidade Federal de Uberlândia, IFGoiano)*; Flavio Silva (Universidade Federal de Uberlândia)

rodrigofr@gmail.com; flavio@ufu.br

Resumo: The Internet of Things (IoT) senses the real world in different domains of knowledge. In this are, in the same realm such as an industry or a city, different IoT based systems can provide several data. In general, these data come from various data sources and the ability to process them in an integrated way would offer new insights and knowledge. The technologies of Cloud can hinder the integration of systems and IoT applications and bring difficulties of interoperability because it has technical dependences involving interfaces and security. This work objective is to tackle these issues by proposing an Open-and-Secure Data Framework (OSDFramework), which is an extension of the open-data concept. OSDFramework aims to create a repository with data from different IoT applications deployed in Cloud environments and is designed to: receive data from applications interested in sharing data; integrate semantically the data collected for sharing, and share data with new systems created. For this, the framework has a data management module, which involves transformations of patterns and models of source data to a common standard, and the security module, which consists of a specification and use of security standards in the different parts. The empirical research method is adopted in this work. The implementation of OSDFramework depends on research activities involving its components. Research is currently being done to raise the possibilities to use of patterns and data models. It is intended to empirically evaluate and make comparisons so that the best decisions are made for the project. The proposal encourages research directions, which relate to aspects such as semantic data interoperability, privacy, and security in computer systems. It is necessary to investigate data management in the Cloud to make the repository viable, which involves process security and the computational cost for the actual application. Data standardization is an essential design problem, as there is a need for an algorithm to transform data from source patterns to a common standard, which is related to interoperability and privacy. Finally, the challenges of communication and security between the parties are presented, considering the ease necessary for the parties to connect with the repository, which depends on a deepening of possibilities and security services, such as protocols, encryption, access control, a method for a consensus between the parties, and possible influences on architecture.

Trilha: Trabalhos de Doutorado

*Apresentador(a) do trabalho.

Classificação de tumores de mama e colorretais baseada em percolação de imagens normalizadas

Guilherme Freire Roberto (Universidade Federal de Uberlândia)*; Marcelo Zanchetta do Nascimento (Universidade Federal de Uberlândia); Thaína Tosta (Universidade Federal do ABC); Leandro Alves Neves (Universidade de São Paulo)

guilhermefroberto@gmail.com; marcelo.zanchetta@gmail.com; tosta.thaina@gmail.com; neves.leandro@gmail.com

Resumo: Percolação é um descritor fractal que tem sido aplicado recentemente em problemas de visão computacional. Esse descritor foi aplicado em 58 imagens histológicas de mama e 165 imagens histológicas colorretais, ambas coloridas com os corantes Hematoxilina e Eosina (H&E), com o objetivo de extrair atributos para diferenciar casos malignos e benignos. Os experimentos também foram realizados sobre imagens normalizadas, visando analisar a influência de diferentes técnicas de normalização de cor em atributos baseados em percolação e se estes são capazes de fornecer melhores resultados de classificação. Os conjuntos de atributos obtidos a partir da aplicação do método em imagens originais e em imagens normalizadas com três diferentes técnicas foram testadas usando 12 classificadores diferentes. Os resultados obtidos foram comparados com outros métodos relevantes da área e contribuições significativas foram observadas, com valores de área sob a curva ROC (AUC) acima de 0,900, tanto em imagens normalizadas quanto em não-normalizadas. Também verificou-se que a normalização de cores não contribui para a classificação de tumores de mama, quando associados à atributos de percolação. Entretanto, imagens normalizadas do conjunto de tumores colorretais forneceram resultados melhores que as imagens originais.

Trilha: Trabalhos de Doutorado

*Apresentador(a) do trabalho.

Geração de documentação a partir de consultas a serviços de busca

Adriano Rocha (Universidade Federal de Uberlândia)*; Marcelo Maia (Universidade Federal de Uberlândia)

adriano.comp2@gmail.com; marcelo.almeida.maia@gmail.com

Resumo: Um problema enfrentado por desenvolvedores de software ao aprender novas tecnologias é a carência de exemplos práticos em suas documentações oficiais. Vários trabalhos têm sido propostos com objetivo de minimizar esse problema por meio de geração de documentações alternativas, porém, esses trabalhos ainda possuem limitações. Por exemplo, as documentações geradas são extensas, o que dificulta a aprendizagem da tecnologia por parte do desenvolvedor. Outra alternativa adotada por desenvolvedores é recorrer a Internet em busca de exemplos práticos. Estes frequentemente utilizam serviços de buscas, tais como: Google, Yahoo, Bing, etc., mas nem sempre as melhores soluções estão entre as primeiras páginas do ranqueamento retornado. Dentre as primeiras retornadas, existem páginas: sem exemplos práticos, com soluções defasadas ou incomuns, sem foco, sem aderência e soluções extensas. Para superar esses problemas, este trabalho tem como objetivo a geração de documentação baseada em consultas relacionadas a atividades de desenvolvimento de software feita por desenvolvedores em serviços de busca. Na primeira etapa da metodologia proposta o desenvolvedor pesquisará uma atividade relacionada ao desenvolvimento de software, por exemplo, “How to make login with php mysql” em um serviço de busca (Google, Yahoo, Bing, etc.). As primeiras páginas retornadas são processadas com o objetivo de selecionar a melhores soluções, a partir das quais é gerada a documentação. Como resultado preliminar, foram realizadas avaliações em 4 pesquisas relacionadas a atividades de desenvolvimento de software utilizando Google, onde foram consideradas as 20 primeiras páginas retornadas pelo serviço de busca para cada uma das pesquisas (80 páginas no total). Para avaliar as páginas retornadas, foram definidos 6 critérios: defasadas/incomuns, sem exemplos práticos, sem foco, sem aderência, soluções extensas e soluções simples/intuitivas. Os resultados mostram que menos da metade (38,75%) das páginas retornadas possuem soluções simples e intuitivas. O restante das soluções (61,25%) não é interessante para os desenvolvedores. Outro ponto importante é que muitas páginas retornadas (13,75%) não possuem exemplos práticos que ensinam como implementar a atividade pesquisada. Também existem páginas (16,25%) com soluções pouco utilizadas ou até mesmo defasadas. Como foi observado nos resultados preliminares, as páginas com soluções simples e intuitivas estão misturadas com outras menos atrativas para os desenvolvedores de software. A abordagem propõe filtrar somente soluções simples e intuitivas retornadas pelo serviço de busca para gerar a documentação. A principal vantagem dessa documentação é que seu conteúdo é focado somente nos elementos técnicos relacionados a atividade que o desenvolvedor deseja aprender, diferentemente de outras abordagens que geram documentações para todos os elementos técnicos da tecnologia.

Trilha: Trabalhos de Doutorado

*Apresentador(a) do trabalho.

Learning coarse-grained code structures

Carlos Eduardo Carvalho Dantas (Universidade Federal de Uberlândia)*; Marcelo Maia (Universidade Federal de Uberlândia)

carloveduardoxpto@gmail.com; marcelo.maia@ufu.br

Resumo: Open-source repositories are often employed to detect or mine patterns that characterize systems, whether for bug detection, code completion, constraints on the sequence of methods to be invoked in APIs, and so on. These patterns can be used to build recommendations systems that suggest practices already used in other softwares, offering an alternative to the standard documentation provided by programming languages and frameworks. Several approaches have been employed to detect or mine patterns, from probabilistic models, n-gram language models, to Deep Learning methods. These approaches can use textual representations, in trees or graphs of the respective repositories. However, a concentration of efforts has been observed in detecting or mining patterns at a finer granularity of the source code. One possibility to be investigated is the mining of structural patterns, from the connection between components that make up a software. For example, a refactoring operation to decompose excessively large classes can involve several components of a software, and its possible to mine the patterns of structural modifications that have already been made in other softwares. These can be useful to developers, helping to design softwares based on commonly used structural patterns. For this, it is assumed as hypothesis that the techniques used until then could also be used to detect these structural patterns.

Trilha: Trabalhos de Doutorado

*Apresentador(a) do trabalho.