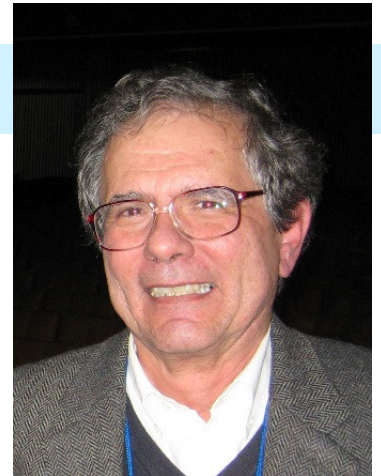


Entrevista com Marco Antonio Casanova Prêmio SBC Mérito Científico 2012

Entrevista por Fernanda Baião, fernanda.baião@uniriotec.br, UNIRIO.

Nesta edição da Coluna Perfil, temos o prazer de conhecer um pouco mais sobre o Professor Marco Antonio Casanova (PUC-Rio), que foi recentemente agraciado com o Prêmio Mérito Científico 2012. Esse prêmio destina-se ao sócio efetivo ou fundador da SBC com reconhecida contribuição científica e/ou técnica em uma das várias áreas e especialidades da computação abrangidas pela Sociedade.



Conte-nos um pouco sobre como você recebeu o Prêmio Mérito Científico de 2012.

Eu diria que foi a estória do “SS – Surpresa e Satisfação”. Surpresa porque eu realmente não esperava e nem sabia que estava sendo cogitado ao prêmio, e satisfação porque, a esta altura do campeonato, a gente espera um empurrãozinho, e é sempre interessante ver que as pessoas lembram do que você fez há 30 anos. E tudo o que eu fiz foi com um monte de co-autores, dentre eles o Furtado é certamente a pessoa com quem eu mais escrevi trabalhos. Estes trabalhos nascem como ideias informais e depois a gente passa muito tempo tornando estas ideias em algo muito claro para os leitores. E muitas das coisas das quais reclamamos, como deadlines e limites de páginas, no final das contas são úteis, pois se não houvesse deadlines a gente nunca parava de escrever um paper, e sem limite de páginas a gente tenderia a escrever muita coisa talvez até desnecessária, que torna a leitura mais enfadonha, e é bom a gente ser conciso. Portanto, “Colabore com seus amigos, não se esqueça dos deadlines e procure ser conciso”.

O que você pessoalmente destacaria dentre suas várias contribuições científicas ao longo de toda a sua carreira?

Tem alguns papers que a gente sempre se lembra. Um paper meu bastante referenciado é um de 1979. Na época toda a teoria de banco de dados presumia a existência da “relação universal” (da qual todas as relações de um banco de dados eram projeções), o que era completamente fora de propósito, mas necessário para toda a teoria de banco de dados funcionar. Neste paper propúnhamos uma nova teoria que envolvia dependências funcionais e dependências de inclusão (atualmente conhecidas como chaves estrangeiras), e mostrávamos um exemplo muito simples que evidenciava que implicação lógica finita e implicação lógica irrestrita eram duas coisas distintas. Foi um dos primeiros casos onde apareceu a noção de implicação lógica finitária em bancos de dados. O interessante é que, apesar de começar por coisas muito simples, o paper mostrou algumas surpresas ao longo do caminho, porque a teoria vai tornando as coisas mais complexas e mais interessantes.

Outra publicação que eu me lembro foi a que me deu mais trabalho até hoje, que foi o livro de programação em lógica. Na época, costumava-se dizer que Prolog era baseado em cláusulas e usava resolução como método de refutação, mas com um estudo mais detalhado você percebe que Prolog trabalha com cláusulas definidas (que são um caso particular de cláusulas) e que o sistema de refutação por trás do Prolog é eliminação de modelos, que não é exatamente igual à resolução. Neste livro, nós começamos a pensar e discutir como seria se começássemos a usar cláusulas genéricas e qual era o grande diferencial entre programação em lógica e prova automática de teoremas. E então pensamos de 3 maneiras diferentes. O primeiro ponto era que, quando se trabalha com cláusulas definidas, o conjunto de cláusulas define um modelo mínimo; no entanto, com cláusulas genéricas não há mais esta propriedade. O segundo ponto que mostramos foi que era possível estender a máquina do Prolog para cláusulas genéricas, e de uma maneira extremamente simples, bastando para isso permitir que a máquina de interpretação do Prolog não olhasse apenas para o topo da pilha durante a criação da árvore

Esta é uma publicação eletrônica da Sociedade Brasileira de Computação – SBC. Qualquer opinião pessoal não pode ser atribuída como da SBC. A responsabilidade sobre o seu conteúdo e a sua autoria é inteiramente dos autores de cada artigo.



Prof. Casanova, na plateia do CSBC 2012 antes da premiação

de refutação, mas também para dentro da pilha. O terceiro ponto foi mais complicado, que foi demonstrar que dava para estender a noção de resposta (ou seja, de forma breve, qual substituição das variáveis torna a consulta consequência lógica dos axiomas) para cláusulas genéricas. A prova disso deu muito trabalho, por conta dos vários detalhes que tiveram que ser tratados. Na época quem trabalhava nisso era uma aluna minha de mestrado, Andrea Silva Hemerly, que ganhou o prêmio de melhor dissertação daquele ano.

Um terceiro trabalho que eu citaria surgiu mais recentemente, quando em minha palestra do SBBB 2007, em João Pessoa, eu comentei que se deve projetar o esquema externo do banco de dados reusando ontologias. Isso levantou uma questão porque, de fato, não se reusa uma ontologia, mas sim um “pedaço” dela, ou seja, um subconjunto de seus termos (vocabulário) e mais as restrições de integridade que se aplicam a estes termos. A questão

interessante era exatamente como computar e exibir para o usuário este conjunto de restrições de integridade (que se aplicam aos termos selecionados). Isto vai além de simplesmente saber se uma afirmação é consequência lógica de outras. Isso deu origem ao artigo do DKE Journal¹, onde propusemos o arcabouço para resolver esta questão. Na minha palestra do AMW 2012, em Ouro Preto, eu mostrei uma álgebra de ontologias para reuso de construções conceituais. Este foi um bom exemplo onde a resposta a uma pergunta simples não é tão simples assim. Você esbarra em limitações e tem que assumir algumas restrições, e é bom que estas restrições façam sentido.

Você tem uma palestra muito interessante disponível em seu Web site, que fala sobre “Bancos de Dados: De onde vieram? Onde estão? Para onde vão?”. Como ganhador do Prêmio Mérito Científico de 2012, poderia nos falar um pouco sobre estas visões, trazidas para os dias de hoje?

Deixa eu comentar um pouquinho sobre esta palestra. Nela há duas fotografias, a primeira é uma obra-prima de Paul Gauguin², que é exatamente sobre de onde viemos, onde estamos e para onde vamos, e a segunda é um mural de Cícero Dias³. Bem, eu levei uns 30 segundos para encontrar a primeira fotografia usando o Google, simplesmente usando as palavras certas porque eu tinha algumas dicas sobre ela. Já a fotografia do mural do Cícero Dias eu levei bastante mais tempo para encontrar, porque eu já a tinha visto em algum lugar que não me lembrava onde, então não tinha boas pistas. Na palestra, eu comento como isto é um bom uso da Web, discuto o que isto tem diferente de um banco de dados, e por aí afora. A palestra então fala da história dos bancos de dados e de dois marcos importantes, que são os recipientes do Turing Award: o Codd⁴, que foi o primeiro da área de Bancos de Dados a receber, e depois o Jim Gray⁵. Depois eu menciono a série de grandes conferências de bancos de dados, Asilomar, e mais recentemente eu cito esta ideia do “Big Data”, que traz grandes problemas. Mas não podemos também esquecer que há também o “Small Data”, os bancos de dados que estão replicados em vários lugares diferentes (por exemplo, os dados da sua agenda ou contatos que estão no seu celular, no seu iPad, no seu computador) e precisam ser sincronizados de uma maneira descomplicada. A gente tem aí dois problemas diferentes: um é a usabilidade, porque as pessoas comuns estão cada vez com mais poder computacional na mão, e você tem que fazer as coisas de maneira descomplicada. Já o “Big Data” é o outro extremo, pois de algum tempo para cá há coleções de dados muito grandes; há um exemplo de aplicação no LNCC, um catálogo de objetos de astronomia, com uma tabela com 1 bilhão de entradas e milhares de atributos, e esses dados ainda são versionados. Com isso, as pessoas começam a dizer que tudo que se fez em bancos de dados até hoje não serve mais, o que é uma absoluta mentira. Os bancos de dados que a gente conhece até hoje vão continuar existindo, ninguém vai se preocupar em construir um sistema de reserva de passagens sem um controle de concorrência; por outro lado, para um catálogo de estrelas este não é um aspecto importante, mas versionamento e identificação dos objetos são. Alguns dos problemas que são tratados pelos SGBDs de hoje (noções de

transação, recuperação e outros) não são relevantes para estas coleções de dados gigantescas, mas outros são. Para o cenário do Small Data, o antigo problema de replicação de dados continua lá, agora em uma escala pessoal, e é uma questão muito séria que a gente precisa resolver.

Então, para onde vão os Bancos de Dados? Hoje em dia a gente tem mais bancos de dados com mais aplicações do que qualquer



Presidente da SBC Paulo Cunha, Prof. Nívio Ziviani (Prêmio Mérito Científico 2011) e Prof. Casanova

outra coisa que a gente já teve no passado. E cada um destes nichos de aplicação tem as suas necessidades, em que parte delas são atendidas pelos SGBDs existentes, e para parte delas os SGBDs existentes são muito complicados, e requerem coisas mais interessantes. Isso é bom, porque há 10 anos não havia interesse por Bancos de Dados pois parecia que todos os problemas estavam resolvidos. E de repente aparece uma coleção enorme de outros problemas que precisam ser tratados.

Se você tivesse que projetar estas mesmas reflexões sobre a Computação em geral, o que destacaria?

Ontem eu atualizei a máquina Java no meu computador e recebi um aviso assustador, dizendo que Java roda em mais de 3 bilhões de aparelhos, incluindo vários equipamentos diferentes. Ou seja, existem 3 bilhões de “coisas” por aí que têm poder computacional (e rodam Java). Se pensarmos que a humanidade tem 9 bilhões de pessoas, quase 1/3 da população da humanidade é o número de objetos que rodam Java. Isso significa que as coisas têm que ser simples de usar. Você não pode imaginar uma televisão que precise se conectar com a Internet e esperar 2 horas para baixar uma nova versão daquele software; isto acontece hoje e é um problema que precisa ser resolvido. Outro problema que permeia tudo em computação é a usabilidade, pois vários aparelhos hoje dependem de computação tanto quanto de qualquer outra coisa que se fez até hoje. Também por isso a confiabilidade é outro problema muito sério, não dá mais para imaginar hoje em dia a gente usar alguma coisa com poder computacional que de vez em quando para no meio do caminho. Então, a gente tem a responsabilidade de fazer coisas fáceis de usar e que sejam confiáveis. O terceiro problema que é bem sério é a capacidade que a gente tem de manter as coisas. Será que a gente daqui a 10 anos vai ter a capacidade de manter as coisas que criamos hoje? A gente tem que ter condições de motivar a próxima geração da computação a entender como as coisas funcionam e mantê-las. Pelos menos estes 3 são meus desafios de pesquisa atualmente.

Como você descreveria sua trajetória profissional e envolvimento com a SBC?

Eu sou um sócio bem antigo da SBC e me lembro claramente de 2 situações: uma quando eu trabalhava na IBM e outra já aqui no Departamento de Informática da PUC-Rio, em que na época eu fiquei espantado de não serem sócios institucionais da SBC. Depois eles se associaram. Eu fico preocupado com que as pessoas se interessem em ser sócios da SBC, e as “pessoas” institucionais também, onde houver sentido. Tem que participar da nossa comunidade, que é quem influencia a Computação no Brasil. Imagino que a SBC realize campanhas junto a pessoas jurídicas e junto aos alunos para se associarem.

A nossa revista é dedicada a estudantes e jovens profissionais. Como você vê o nível de preparação atual de nossos estudantes de graduação e pós-graduação na área de computação e as possibilidades de atuação desses profissionais? Que conselho você dividiria com nossos leitores?

A possibilidade de atuação de quem trabalha em Computação é total, desde trabalhar junto a físicos próximo de um acelerador atômico até trabalhar criando aplicativos para telefones. Uma vez que a

Computação hoje permeia literalmente tudo, a vantagem é que quando um setor da economia está mal, outro vai bem, então o profissional (da Computação) provavelmente estará bem de emprego sempre. A Computação é algo que sempre tem futuro e sempre vai ter, cada vez mais. A razão de hoje termos tão poucas pessoas trabalhando em Computação quando comparado há tempos atrás talvez seja porque a Computação exija muito esforço, do ponto de vista de lidar com complexidade. Você precisa ter uma formação um pouco mais voltada para as ciências exatas. Por exemplo, o rigor que é exigido em algumas disciplinas como Cálculo ajuda a acostumar (os alunos) com formalismo, e o rigor que se aprende em Lógica é essencial para você criar artefatos de software bem feitos.

Computação é uma boa área para se trabalhar, sempre haverá emprego. Mas o aluno tem que ter paciência em estudar disciplinas mais formais, que lhe deem uma formação mais precisa, e tem que estar preparado para lidar com abstrações e detalhes, pois Computação é muito detalhista e, se você omite os detalhes, deixa de lado a questão da qualidade, que é fundamental. A recompensa é que, acredito que por muito tempo ainda, o seu emprego estará garantido.

A seguir, apresentamos na íntegra o Discurso por ocasião da cerimônia de entrega do Prêmio Mérito Científico da SBC (transcrito e editado pelo autor).

Inicialmente gostaria de agradecer aos meus colegas pela indicação, que recebi com surpresa e satisfação. Naturalmente, gostaria de agradecer aos meus inúmeros coautores, alguns dos quais estão aqui presentes. Em particular, gostaria de agradecer ao Prof. Furtado, que não está presente, e com quem escrevi bem mais da metade dos meus trabalhos.

Tinha algumas opções para este discurso de agradecimento. Resumir as contribuições ao longo da carreira seria provavelmente a pior delas. Sem ajuda de um projetor seria pior ainda. A segunda opção seria fazer um paralelo com a obra prima do Ingmar Bergman, “Morangos Silvestres”, onde um velho professor está a caminho da sua *alma mater* para receber um reconhecimento e tem pesadelos recorrentes em que é reprovado em exames por não se lembrar de mais nada da profissão. Isto certamente é o meu caso. Na verdade, diplomas de doutorado deveriam ter uma data validade. Porém, já desisti de usar os meus ícones culturais depois que repetidamente escutei dos alunos... “mas eu não era nem nascido!” (e, portanto, não sabia do que eu estava falando). A terceira opção seria contar para vocês, metaforicamente, como se escreve um paper.

Então vamos colocar a seguinte pergunta: “Com quantos cafezinhos se escreve um paper?” Eu diria que é da ordem de 100 cafezinhos. Portanto, quem tiver a curiosidade de olhar o meu currículo, vai rapidamente perceber que ele equivale a uns 20.000 cafezinhos.

O problema é que a metáfora é quase real. Hoje em dia, cedo pela manhã, tomo um cafezinho com a Karin (Breitman, profa. Puc-Rio), onde 5 – 10 minutos de conversa são suficientes para começar a alinhar uma ideia. No meio da manhã e no meio da tarde, quando as aulas permitem, tomamos outros cafezinhos com o Furtado. A conversa gira em torno do que andamos escrevendo. Assim, dos 100 cafezinhos para escrever um paper, uns 10 – ou seja, 2 ou 3 semanas – são para alinhar as ideias. Correspondem ao trabalho cooperativo e criativo que cristaliza as ideias até ao ponto em que podem ser colocadas no papel. Os outros 90 cafezinhos são para escrever o paper. O processo consiste em escrever, tomar cafezinho para discutir, reescrever, tomar cafezinho para rediscutir, e assim vai por uns 2 a 3 meses, o que corresponde então aos outros 90 cafezinhos.

Ou seja, depois que você tem as suas ideias mais ou menos consolidadas, o processo de escrever consiste em explicá-las da maneira mais clara possível: definições, lemas, teoremas, algoritmos, análises de desempenho e por ai vai. O texto tem de ser tal que nenhum dos coautores tenha qualquer dúvida a respeito do que está escrito. A única maneira que conheço para isto está em escrever, ler, discutir e reescrever, até que cheguem o famoso deadline e o não menos famoso limite de páginas.

Deadlines são muito importantes neste processo. Caso contrário, você cai em uma reescrita sem fim, com ganhos (de clareza) cada vez menores. O deadline corta a cauda deste processo. O limite de páginas é duplamente importante: para deixar alguma coisa para o próximo paper e para forçar você a exercitar a sua capacidade de sintetizar ideias. Frequentemente você vai descobrir parágrafos redundantes ou com divagações que só atrapalham a leitura do texto.

O que significa tudo isto, enfim? Um professor de medicina, se existir algum na plateia, vai imaginar um estudo com este grupo de pessoas (os coautores dos papers) para estabelecer o efeito da cafeína na atividade dos neurônios. Um professor de psicologia, da linha behaviorista, iria imaginar outro estudo, mostrando que é possível condicionar um grupo de pessoas a tomar cafezinho e escrever papers. Se isto fosse verdade, o problema da produção de um programa de pós-graduação estaria resolvido: bastaria condicionar os alunos de doutorado, durante seis meses a um ano, dando cafezinho e ensinando a escrever papers; no resto do doutorado, bastaria dar cafezinho para que os papers saiam automaticamente. É claro que não era este o meu ponto: este foi apenas um desvio para distrair vocês (a plateia).

O que quis dizer então com a metáfora do cafezinho? Ela é uma metáfora, primeiro, para trabalho colaborativo, desinteressado e criativo, algo que aprendi com o Furtado. Nós nunca discutimos quem seria o primeiro autor: seria aquele que não o foi no paper anterior. Eu procuro transmitir isto para os coautores mais jovens. Na segunda parte da metáfora, chamo a atenção para o contínuo aperfeiçoamento do texto que se segue: escrever, ler e reescrever até que as ideias estejam expostas de forma clara e rigorosa. Também recomendo não reclamar de limites. Limites de tempo e de páginas são desafios que nos forçam a escrever papers mais concisos e mais claros.

Se vocês seguirem esta receita durante uns trinta e poucos anos, talvez vocês cheguem até aqui (este pódio), com a ajuda dos amigos. É isto que gostaria de deixar com vocês: cooperação, coautores, cafezinhos, deadlines e limites de páginas.

Obrigado a todos.

Notas do Autor

¹ Marco A. Casanova, Tanara Lauschner, Luiz André P. Paes Leme, Karin Koogan Breitman, Antonio L. Furtado, Vânia Maria Ponte Vidal: Revising the constraints of lightweight mediated schemas. *Data Knowl. Eng.* 69(12): 1274-1301 (2010)

² Paul Gauguin, French (1848–1903), *Where Do We Come From? What Are We? Where Are We Going?*, 1897–1898, Oil on canvas, Wildenstein 561, Museum of Fine Arts, Boston: Tompkins Collection 36.270, <http://www.mfa.org/artemis/fullrecord.asp?oid=32558&did=500>

³ Cicero Dias, Brasil, 2003, *Eu vi o mundo...* Ele começava no Recife, Rio de Janeiro, 1926-1929, Coleção do artista, Paris, <http://www.estadao.com.br/divirtaseonline/galeria/cicero dias/painel/index.frm>

⁴ http://amturing.acm.org/award_winners/codd_1000892.cfm

⁵ http://amturing.acm.org/award_winners/gray_3649936.cfm