

Modelagem de Sistemas Computacionais Complexos: Em Direção a uma Fundamentação Científica e Aplicações Tecnológicas*

Luís C. Lamb¹, Aline Villavicencio¹, Marco Idiart², Luciana S. Buriol¹, Leila Ribeiro¹, Roberto da Silva¹,

¹Instituto de Informática, ²Instituto de Física – Univ. Federal do Rio Grande do Sul
91501-970 – Porto Alegre, RS – Brazil
{lamb,avillavicencio,leila,buriol,rdasilva}@inf.ufrgs.br
idiart@if.ufrgs.br

Abstract. *We present a long term view on scientific and technological challenges in complex systems development. We raise several research hypotheses: (1) the foundations of complex systems are built upon computer science, physics and natural sciences, linguistics, cognitive science, and economics; (2) the development of sound models of complex computational systems leading to technological advances demands researchers with in-depth knowledge of such disciplines; (3) the construction of complex systems leading to relevant technologies requires interdisciplinary research and new ways of soundly combining research results from several disciplines. We briefly describe five related research paths, outlining challenges that may well lead to the construction of large-scale, cognitive and biologically-inspired complex computational systems impacting on social, cognitive and physical sciences.*

Resumo. *Apresentamos uma visão de longo prazo em desafios científicos e tecnológicos no desenvolvimento de sistemas complexos. Levantamos diversas hipóteses de pesquisa, entre elas: (1) a fundamentação de sistemas complexos é construída sobre sólida base de ciência da computação, física e ciências naturais, lingüística, ciência cognitiva e economia; (2) o desenvolvimento de modelos corretos de sistemas complexos computacionais conduzindo a avanços tecnológicos demanda conhecimento aprofundado destas disciplinas; (3) o desenvolvimento de modelos de sistemas computacionais complexos conduzindo a avanços tecnológicos requer pesquisa interdisciplinar e novas formas de corretamente combinar resultados de pesquisa de diversas disciplinas. Descrevemos brevemente cinco linhas de pesquisa relacionadas, destacando desafios que podem levar à construção de sistemas computacionais complexos de ampla escala, cognitiva e biologicamente inspirados, com impacto nas ciências físicas, cognitivas e sociais.*

1. Introdução

Historicamente, a ciência da computação construiu sua fundamentação sobre conceitos lógico-matemáticos. O impacto desta fundamentação é conhecimento comum, pois tecnologias e metodologias, e.g, em engenharia de software, bancos de dados, sistemas distribuídos, inteligência artificial, desenho e projeto de hardware são fortemente baseados em construções lógico-matemáticas. Assim como E. Wigner, vencedor do

* Pesquisa parcialmente financiada pelo CNPq.

Nobel de Física em 1963, defendeu e demonstrou que o papel da matemática é fundamental na modelagem de fenômenos físicos [Wigner 1961], em computação a lógica e matemática têm historicamente desempenhado papel semelhante [Halpern 2001]. Cabe salientar, que em *loc.cit.* Halpern et al ilustram as aplicações de lógica em bancos de dados, linguagens de programação, complexidade descritiva, verificação de circuitos digitais, inteligência artificial e sistemas distribuídos.

A evolução da computação tem levado a implicações e impactos sociais crescentes, como a possibilidade de interação e integração globais, tanto social, quanto econômica. Além disso, a crescente interação entre a computação e as ciências naturais, humanas e sociais tem levado à construção de sistemas cada vez mais sofisticados, que exigem de pesquisadores e profissionais tanto na academia, quanto na indústria, de conhecimentos multi-disciplinares. Entre estes sistemas, destacam-se os *sistemas complexos computacionais*. O termo sistema complexo é bastante abrangente e se origina da mecânica estatística. Historicamente, Boltzmann (1844-1906) foi pioneiro ao descrever formalmente a distribuição de probabilidades da energia de um gás. Boltzmann mostrou que o comportamento da energia média de um gás relativamente simples era essencialmente a sua temperatura, mostrando talvez pela primeira vez essa a possibilidade *emergência*, conceito amplamente utilizado em computação. Existem muitas definições de complexidade. Por esta razão muitos objetos naturais, artificiais e abstratos ou mesmo redes podem ser considerados sistemas complexos.

A idéia de que um sistema com muitos componentes com regras e características locais possa fazer emergir regras gerais e muitas vezes universais culminou na chamada teoria dos fenômenos críticos e transições de fase [H. E. Stanley, 1971]. Por outro lado, [Metropolis et al, 1953] incorporou a distribuição de Boltzmann e desenvolveu um algoritmo que obtém as propriedades termodinâmicas e, portanto, globais de um gás usando um processo conhecido como amostragem por importância. Este algoritmo (de Metropolis), um dos 10 algoritmos mais utilizados no século, abrange aplicações em engenharia, física e computação. O algoritmo de Metropolis consiste em uma importante parte de um outro algoritmo extremamente útil: *Simulated Annealing*, uma heurística eficiente para determinação de mínimos locais próximos a mínimos globais em sistemas desordenados. Em um cenário de ciência de computação, uma aplicação deste algoritmo ocorre no multidisciplinar problema do *caixeiro viajante*. A idéia destes algoritmos tipo *Monte Carlo*, culminou nos chamados algoritmos genéticos. A incorporação da aprendizagem, nas ditas redes neurais levou ao surgimento de uma teoria mais geral em ciência da computação denominada *Machine Learning*.

Respondendo aos desafios de construir sistemas complexos computacionais, este artigo levanta possibilidades de pesquisa e propõe desafios e requisitos para a construção destes sistemas. Estes sistemas apresentam, por exemplo, dimensões que modelam aprendizagem, raciocínio, distribuição, *multi-agência*, plausibilidade biológica, rigorismo formal e precisão matemática. Além disso, o seu desenvolvimento exige conhecimento multi e transdisciplinar, pois estes sistemas devem interagir e estar integrados a ambientes com forte participação do homem. A seguir, enumeramos desafios, requisitos e características desejáveis de cinco tipos de sistemas computacionais complexos e seu potencial impacto no desenvolvimento da computação.

2. Desafios no Desenvolvimento de Sistemas Complexos Cognitivos

A computação evoluiu e surgiu de noções fundamentais em lógica e matemática. Não-especialistas associam o desenvolvimento da computação à construção de mecanismos

tecnológicos de cálculo e aritmética, percebendo a consequência do conhecimento científico fundamental. Esta situação persiste até hoje, pois, exceto entre especialistas, o desenvolvimento científico somente é visto como relevante pelo grande público se a consequência de pesquisas básicas de longo prazo conduzem a resultados de impacto como, e.g., o desenvolvimento do transistor, o surgimento de computadores, a ubiquidade dos mecanismos de comunicação e o surgimento da web, para nos restringirmos apenas à computação.

Algumas vezes, o desenvolvimento e pesquisa básica em ciência conduzem à quebra de paradigmas, como, e.g., o resultante dos desenvolvimentos fundamentais em mecânica quântica (de Planck, Dirac e muitos outros), da eletricidade e magnetismo (Maxwell, Faraday e outros), e da própria mecânica Newtoniana, por sua vez fortemente baseada em desenvolvimentos em análise (e por consequência, no cálculo). Na computação, a maior quebra de paradigma foi a própria construção de computadores nas décadas de 1930 a 1950, fortemente influenciados por desenvolvimentos em lógica matemática a partir dos trabalhos de Turing, Gödel e muitos outros. Esta quebra de paradigma foi tão marcante que influenciou modelos micro e macroeconômicos nas décadas seguintes. Outra quebra de paradigma aconteceu recentemente com a criação da web, que permitiu integração e interação rápida entre usuários e posteriormente entre (potencialmente) toda população. Também alterou radicalmente o modelo de cooperação e desenvolvimento científico, permitindo integração e acesso rápido ao conhecimento não vista anteriormente em outros períodos da história. Nota-se que a atual quebra de paradigma, ao contrário das anteriores, é fortemente baseada em modelos e dimensões cognitivas, intrínsecos não somente ao indivíduo, mas também a grupos e largas populações, exigindo análise social do comportamento de grupos.

Modelos de computação para sistemas complexos que contemplem cognição devem então observar e responder à várias dimensões, individuais e coletivas, e interagir com o mundo exterior. A construção de sistemas complexos cognitivos, considerando habilidades integradas de percepção, comunicação e cooperação são ainda incipientes. Para responder a este desafio, é necessário dominar e compreender os modelos da cognição. A ciência cognitiva é construída sobre a psicologia, filosofia, lingüística, neurociência, antropologia, computação e inteligência artificial. Portanto, modelos e sistemas que tenham como objetivo modelar a interação homem-máquina com eficiência, adaptabilidade e usabilidade devem considerar aspectos cognitivos. Devemos então modelar aspectos internos e externos ao sistema, como adaptabilidade e aprendizagem, abstração lógico-matemática adequada e mecanismos de raciocínio computacionalmente tratáveis.

Valiant, em 2003 (em artigo visionário sobre os próximos 50 anos da computação, publicado no *Journal of the ACM*, contendo artigos dos ganhadores do *Nevanlinna Prize* e *A.M. Turing Award* [Valiant 2003]), defendeu que o desenvolvimento de uma *semântica para a computação cognitiva* é um dos maiores desafios da ciência da computação, juntamente com a caracterização da computação cortical. Para responder a estes desafios é essencial entender os mecanismos de representação de conhecimento, aprendizagem e raciocínio no cérebro humano. Neste sentido, diversos autores têm desenvolvido modelos de computação cognitiva que tentem explicar mecanismos de representação do cérebro [Hinton 2005], integrem raciocínio e aprendizagem [Garcez e Lamb, 2006; Richardson e Domingos 2006]; e que modelem lógica, física ou matematicamente a cognição [Lamb, Borges e Garcez 2007]. [Thagard 1996] apresenta uma introdução ao problema.

A computação cognitiva e o desafio de construir modelos para o raciocínio e aprendizagem, biologicamente plausíveis e matematicamente tratáveis constitui-se em um desafio de grande dificuldade em sistemas computacionais complexos. Acrescente-se a isto, o fato de que normalmente a aprendizagem de máquina e adaptabilidade têm sido estudadas através de modelagem probabilística. No entanto, mecanismos de modelagem e automação do raciocínio são pesquisados através de mecanismos lógicos. Assim, conciliar a incerteza intrínseca aos modelos probabilísticos de aprendizagem com a precisão dos mecanismos simbólicos, baseados em lógica, constitui-se em um desafio fundamental a ser vencido. Antes disso, a construção de sistemas computacionais complexos cognitivos, que integrem raciocínio e aprendizagem, com benefícios tecnológicos, permanece como assunto de pesquisas em andamento.

A relevância da construção de modelos da mente e cérebro, também tem sido destacada e reconhecida historicamente por muitos. Por exemplo, um dos grandes desafios propostos pela *British Computer Society* é justamente a construção de modelos e arquiteturas para a mente e o cérebro [Hoare e Milner, 2004]. Entre os desafios enumerados, na construção e modelagem do cérebro e mente, estão: a percepção dos tipos de objetos em torno de nós mesmos; a compreensão da linguagem; a aprendizagem de novos conceitos; o controle de nossas ações; a memória dos objetos e percepções; a apreço/desapreço pelas coisas; consciência de nossos pensamentos e emoções; a aprendizagem e reconhecimento dos estados mentais dos outros; o entendimento de músicas e brincadeiras; a percepção da passagem do tempo. Todos estes, de acordo com [Hoare e Milner 2004] envolvem modelos mentais abstratos e processos físicos concretos. O desafio seria construir sistemas complexos com estas características, que nossos cérebros e mentes conseguem com eficiência, mas que modelos e sistemas computacionais até hoje não foram capazes de implementar.

Assim, a construção de modelos computacionais e sistemas cognitivos complexos têm de considerar a intrínseca complexidade do problema e a complexidade do sistema que está se tentando modelar: em última análise, a mente e cérebro humanos. Novos avanços em neurociência, onde são estudados os mecanismos biológicos do cérebro, lingüística e processamento computacional da linguagem, ciência cognitiva e modelos computacionais biologicamente inspirados são essenciais na construção de mecanismos e arquiteturas que expliquem e simulem o comportamento e a computação da mente humana e do cérebro. Além disso, a construção destes modelos cognitivos computacionais ricos que integrem, e.g., raciocínio e aprendizagem tem forte ligação com mecanismos de computação da linguagem natural – um desafio listado na seqüência - uma vez que para interagir e compreender o mundo é essencial que modelos eficazes e computacionalmente tratáveis da linguagem sejam desenvolvidos. O desafio de construir sistemas complexos de cognição é também integrado a modelos estudados em *econofísica*, uma vez que a modelagem de agentes econômicos sob racionalidade limitada exige novos desenvolvimentos em lógica e modelagem da aprendizagem em ambientes distribuídos. Estes ambientes contam com a participação de agentes humanos, devem prover interação homem-máquina eficiente e propiciar a interação homem-natureza. Em última análise, estes sistemas devem auxiliar o homem na sua interação com o mundo, pois, hipoteticamente, representariam com maior fidelidade a cognição humana. Assim, antes de propormos soluções aos desafios enfrentados na construção de sistemas cognitivos complexos, é necessário identificar requisitos necessários para a construção destes tipos de sistemas, que envolvem um grande número de dimensões e áreas de pesquisa.

3. Aquisição e Processamento Computacional da Linguagem: Usando um Melhor Entendimento da Mente Humana na Criação de Sistemas de Computação Adaptativos

Dado que o meio mais natural para o homem comunicar-se é a sua linguagem (Português, Inglês, Alemão,...), pesquisadores da área de Processamento de Linguagem Natural (PLN) têm investigado formas para capturá-la computacionalmente. Um dos principais objetivos da PLN é possibilitar interações tais como aquelas apresentadas no filme “2001: Uma Odisseia no Espaço”, na qual um humano dialoga com um computador (HAL) sem qualquer restrição no uso da linguagem, de ambas as partes. Para tanto, recursos, métodos e ferramentas têm sido desenvolvidos com muito do esforço investigativo sendo empregado na construção de modelos computacionais das linguagens humanas, em particular de como combinar correta sintaxe com pertinência semântica, para uma subsequente seqüência de ações apropriadas à interação. Isto requer não apenas a modelagem de conhecimentos lingüísticos como conhecimentos gerais sobre o mundo.

Apesar dos rápidos progressos da área [Jurafsky e Martin 2000, Manning and Schutze 1999], uma interação como a do filme ainda é um sonho distante. De fato, como propôs o matemático Alan Turing [Turing 1950] com seu celebrado jogo de imitação, a obtenção de um manejo satisfatório de linguagem por parte de uma máquina passa necessariamente pelo entendimento de algo muito mais fundamental: o funcionamento de nossa própria mente consciente. Conseqüentemente a modelagem das linguagens humanas, com toda sua complexidade e dinâmica, é um grande desafio para o conhecimento humano. Desafio este onde a Ciência da Computação certamente contribuirá de forma decisiva.

Há três abordagens principais dentro da área de PLN para modelar estes conhecimentos. A primeira é a mais tradicional, a **baseada em conhecimento** [Jurafski and Martin 2000], que tem influências da Inteligência Artificial, e onde os recursos necessários para o tratamento da linguagem são bases precisas de conhecimento manualmente construídas, ou seja, dicionários e conjuntos de regras compiladas por lingüístas. Porém a construção manual dos recursos lingüísticos necessários é um processo muito lento e oneroso, e resulta em recursos estáticos que são incapazes de se adaptar à constante evolução da linguagem. Uma alternativa para estes problemas é o uso de uma **abordagem estatística** onde métodos estatísticos são utilizados para possibilitar a aquisição automática de informações lingüísticas a partir de uma coleção de textos, por exemplo, visando a construção (semi-)automática de recursos lingüísticos, tais como dicionários e gramáticas [Korhonen 2003, O'Donovan et al. 2004, Zhang et al. 2006]. Porém as técnicas utilizadas em geral necessitam de grandes quantidades de dados que podem não estar disponíveis para todos os domínios e línguas. No contexto de pesquisa do Brasil, várias destas técnicas, e a eventual aplicabilidade das mesmas à Língua Portuguesa, estão sendo desenvolvidas e testadas em trabalho de colaboração entre os institutos de Informática e de Física da UFRGS (descrito em [Zhang et al. 2006]).

A terceira e última abordagem, a qual podemos chamar **abordagem cognitivista**, direciona seu foco não somente aos produtos finais da atividades de linguagem, tais como textos e dicionários, mas principalmente em como a linguagem é adquirida. A aquisição humana da linguagem é um processo fascinante. Uma criança aprende sua língua materna em um tempo curto, bastando estar exposta a ela. Nenhum tipo especial de pedagogia é necessário. Avanços no entendimento deste aspecto da cognição humana

no âmbito de disciplinas como a psicologia, a biologia e a lingüística começaram a informar estas investigações e influenciar os desenvolvimentos na área. Como resultado surgiram novos modelos cognitivos computacionais de processamento da linguagem descrevendo como se daria aquisição, armazenamento e processamento da linguagem humana [Briscoe 1997-1999, Berwick e Niyogi 1996, Sakas e Fodor 2001, Villavicencio 2002]. E estes por sua vez têm sido utilizados em estudos multidisciplinares de aspectos da cognição humana, como o modelo do léxico mental implementado como uma estrutura de grafos, usado na investigação da aquisição humana de significados das palavras [Gaume 2004].

Nos últimos anos também foram coletados e disponibilizados vários corpora (textos) de grande magnitude para diversas Línguas, como o British National Corpus, para a Língua Inglesa, que contém 100 milhões de palavras. Estes corpora possibilitam o estudo sistemático de uma língua e de métodos para a modelagem e aprendizado automático das construções lingüísticas neles contidas. De particular importância para os estudos cognitivos computacionais são os corpora contendo transcrições de gravações com crianças na fase de aquisição da linguagem, em geral obtidos no âmbito de estudos lingüistas e psicolingüistas, sendo a coleção de corpora CHILDES (Child Language Data Exchange System) [MacWhinney 2000] a mais amplamente conhecida. Estes corpora são de grande relevância para estudos que utilizam modelos cognitivos computacionais, pois permitem simular de uma certa maneira o ambiente lingüístico ao qual uma criança é exposta durante o período de aquisição da sua língua materna. Porém muito poucos corpora estão disponíveis para a Língua Portuguesa. A Linguateca (<http://www.linguateca.pt/>) é uma iniciativa interinstitucional para a coleta de textos jornalísticos. Todavia, corpora similares aos do CHILDES para o Português ainda estão em falta, e para que um avanço significativo do PLN seja possível é necessário um esforço conjunto interinstitucional para obter e disponibilizar corpora similares a estes

Encontrar modelos fundamentados em estudos psico-lingüísticos e que ao mesmo tempo sejam computacionalmente implementáveis não só possibilita o avanço na construção de tecnologias adaptativas para PLN, como também permite avanços na outra direção, na utilização destes modelos para se chegar a um melhor entendimento sobre a aquisição, armazenamento e uso da linguagem por humanos. Nas últimas décadas, a aquisição da linguagem tem se tornado o foco central em disciplinas, como a Lingüística e a Psicologia, sendo Noam Chomsky possivelmente o seu mais conhecido contribuinte. Diferentes modelos de aquisição da linguagem [Chomsky 1965, Pinker 1984, Pollard e Sag 1989] etc. e teorias lingüísticas como Government and Binding [Chomsky 1965], Head-Driven Phrase Structure Grammar [Pollard e Sag 1987] e Minimalism [Chomsky 1995] foram propostos para tentar capturar a essência da linguagem humana. Contudo, em geral, os resultados obtidos se limitam apenas a explicações teóricas de fenômenos bem localizados em um número restrito de línguas, muitas vezes idealizando o uso destas línguas. Além disto muitos destes estudos abordam fenômenos dentro de uma disciplina isolada (por exemplo a Psicologia), sem levar em consideração evidências encontradas em outras disciplinas.

Apesar da aquisição da linguagem já fazer parte do domínio de investigação de diversas disciplinas, há ainda muitas questões em aberto. São exemplos: *Quais são as características específicas dos mecanismos e dos processos empregados na aquisição da linguagem? Estes são mecanismos mais gerais de aprendizado, ou são específicos para a linguagem? Antes de ser exposta à uma língua, quanto uma criança sabe sobre linguagens em geral? Seria isto inato ou decorrente do convívio? Existe uma estrutura comum a todas linguagens humanas? Existe uma relação entre esta estrutura e a*

forma particular de processamento existente no cérebro humano?

Já existem muitas propostas para possíveis respostas destas perguntas, sem no entanto uma análise computacional profunda de sua viabilidade como uma explicação de como a aquisição da linguagem ocorre. A ciência da computação dispõe dos recursos teóricos necessários para a realização destes estudos. Através de suas técnicas será possível avaliar e comparar quão bem sucedidos estes modelos seriam na tarefa de adquirir linguagem em tempo limitado, a partir da exposição a um ambiente lingüístico contendo também ruído na forma de sentenças incompletas e não gramaticais, além da possível interferência de outras línguas, etc. Isto é efetivamente o que crianças fazem quando estão aprendendo a sua língua materna. Estudos sobre a capacidade de aprendizado de modelos ajudam a ressaltar potenciais problemas com eles e indicar possíveis caminhos a serem explorados. Porém para que estes modelos possam ajudar a entender melhor o processo de aquisição humana da linguagem, devem empregar somente algoritmos e recursos que um humano possa empregar. Consequentemente, a construção de modelos computacionais de aquisição, armazenamento e uso da linguagem bem fundamentados em estudos psico-lingüísticos e compatíveis com a aquisição humana da linguagem é um dos grandes desafios para a computação. Estes modelos não só auxiliarão no avanço da Ciência da Computação e de suas técnicas de interação homem-máquina através da linguagem natural como também permitirão obter um melhor entendimento sobre a parte da mente humana que lida com processamento de linguagem.

Para finalizar, no que se refere especificamente à Língua Portuguesa, o seu estudo através destas novas perspectivas ainda está em um estágio precário se compararmos com o inglês ou o francês, entre outras. Assim, o Brasil como maior país de Língua Portuguesa e, portanto seu maior usuário, não pode fugir da responsabilidade de acelerar suas pesquisas especificamente em português. Se poder falar com uma máquina é um desafio investigativo da humanidade, poder falar em português com uma máquina é um desafio nosso.

4. Análise e Modelagem do Webgraph

O Webgraph é o grafo formado pela estrutura de links das páginas web, ou seja, as páginas e seus links representam os nós e arcos deste grafo, respectivamente. O Webgraph representa um sistema complexo devido a sua grande dimensão. Em 2005, Gulli e Signorini estimaram que o Webgraph possuía mais de 11.5 bilhões de nós, sendo que atualmente este grafo no mínimo dobrou de tamanho. Em 2004, Boldi e Vigna haviam estimado tal dimensão em 3.5 bilhões nós. Certamente é uma das maiores redes sob pesquisa no mundo e, além disso, é de grande interesse científico e comercial.

Devido à sua grande dimensão, mesmo algoritmos polinomiais representam tarefas árduas devido ao tempo e recursos computacionais requeridos. Além disso, as páginas web são inseridas sem regras, a qualquer momento e por qualquer pessoa, crescendo sem uma organização prédefinida e gerando um grafo sem estrutura conhecida. Por esta razão, muitos dos algoritmos usados para computações no Webgraph são exclusivamente desenvolvidas para ele. Pode-se citar como exemplo as técnicas de compressão que são desenvolvidas especificadamente para comprimir o Webgraph [Boldi e Vigna, 2003; Raghavan e Garcia-Molina, 2003]. Além de ser um grafo de grande dimensão, o Webgraph possui características únicas em relação a outros grafos. A sua expansão e estrutura não apresentam modelos definidos e não segue as mesmas regras comum a outras redes. Mesmo que sua análise estática é de grande

importância, sua transformação no tempo pode adicionar características descritivas essenciais. Conhecer este grafo é de fundamental importância para muitos estudos. Por exemplo, quando se usa uma ferramenta de busca (Google, por exemplo), a ordem em que as páginas resultantes de uma busca aparecem depende significativamente da estrutura deste grafo.

Uma grande dificuldade dos grupos de pesquisa nessa área é a disponibilidade de bases de dados para teste. Aliado a este problema, existe a dificuldade de isolar um sub-Webgraph que mantenha as características originais do mesmo. Além disso, poucas são as bases de dados com informação temporal. Para obter um sub-grafo, define-se qual será o conjunto de nós do grafo mais geral, e então considera-se os links entre nós deste conjunto. Um exemplo de sub-grafo são as páginas web do Brasil, e o grafo resultante seria o Webgraph brasileiro. Para coletar tal subgrafo necessita-se de uma máquina de busca de grande dimensão, banda de rede o suficiente para coleta (atividade esta onerosa em termos financeiros e computacionais), e capacidade de armazenamento e organização dos dados. Ou seja, em geral é uma tarefa realizada entre colaboração de grupos de pesquisa. Outra possibilidade seria isolar um sub-grafo menor, as páginas da UFRGS, por exemplo. Mas quando o sub-conjunto de nós fica muito restrito, a perda dos links externos pode representar a descaracterização do grafo. Bases de dados temporais são praticamente inexistentes. Por exemplo, para ter-se informação da data de criação e datas das atualizações das páginas do Webgrafo brasileiro, seria necessário fazer sucessivas coletas das páginas e comparação do que mudou entre cada par de coletas. O que certamente resultaria em custos onerosos. Ou usar bases de dados alternativas, como é o caso da maior enciclopédia do mundo, Wikipédia [Zambiasi et al., 2007], ou informação de blogs [Kumar et al., 2003].

Finalmente, uma outra possibilidade é a geração de grafos sintéticos, cuja maior dificuldade é definir as regras para geração de grafos que mantenham a estrutura dos Webgraph. Existem algumas propostas conhecidas na literatura, das quais podemos citar *evolving network model* [Albert, Jeon e Barabasi, 1999], *copying model* [Kumar et al., 2000] e *network growth model* [Pennock et al., 2002]. No primeiro modelo, novos arcos são inseridos ligando nós selecionados proporcionalmente ao seu grau de saída, enquanto que no segundo modelo, além da possibilidade usada no primeiro, um arco tem uma probabilidade de ser copiado, mudando apenas o nó origem. Já o terceiro modelo, os links de saída de cada novo nó apontam para nós de acordo com outros fatores, como conectividade e tamanho do conjunto de nós iniciais, além do grau de saída do mesmo. Desta forma, como desafios para estudo de propriedades e modelagem de Webgraph podemos citar: (i) a localização de comunidades dentro Webgraph; (ii) a definição de regras de crescimento do Webgraph. Estes dois itens estão fortemente relacionados. A localização e determinação de comunidades web dentro do Webgraph seria de extrema importância para caracterização deste grafo. Mas os métodos atuais disponíveis permitem determinar apenas comunidades web de pequena dimensão [Kumar et al., 1999]. A detecção de comunidades web é importante porque foi verificado, empiricamente, que os dados das comunidades web são mais atualizados e em geral mais completos. Dessa forma, o ranking de páginas das comunidades web poderiam ser beneficiados e ganhar prioridade.

Um alternativa para o estudo de Webgraphs menores seria se algumas comunidades web amplamente conhecidas apresentarem características semelhantes a grafos muito maiores. É o caso do wikigraph, por exemplo, que é o grafo extraído da estrutura de links dos artigos da enciclopédia online Wikipédia. Estudos recentes

apontam semelhanças entre tais grafos [Zambiasi et al., 2007]. Existe uma facilidade muito maior em estudar um grafo do tamanho da wikipédia (poucos milhões de nós), que o webgraph (bilhões de nós). Em vista do que foi discutido acima, propõe-se a seguinte abordagem para este estudo: (1) extração e estudo de sub-webgraphs que mantenham as características encontradas no webgraph; (2) analisar tais grafos e definir as principais propriedades encontradas; (3) gerar grafos maiores baseados nestas propriedades; (4) verificar se o grafo resultante caracteriza um webgraph.

Embora as tarefas listadas acima sejam claras, o como elas serão realizadas ainda não é claro. Cada um desses itens tem sido um desafio para pesquisadores de diversas áreas, resultando numa gama de publicações nos últimos anos. Diversas conferências e workshops foram recentemente criados para discutir assuntos relacionadas a estes tópicos. Apesar dos avanços, o estudo de propriedades e a modelagem do Webgraph ainda é um desafio para a comunidade científica.

5. Modelagem e Análise Formal de Sistemas Biológicos

Os organismos vivos são sistemas reativos: dependendo de estímulos (condições ambientais e presença ou ausência de certas substâncias), a célula produz uma reação (produzindo as proteínas/enzimas adequadas para tratar o estímulo). A informação sobre quais proteínas/enzimas devem ser geradas para cada situação está guardada no núcleo, no DNA. A área de Biologia Molecular começou a se desenvolver mais intensamente desde o descobrimento da estrutura do DNA, como uma dupla hélice de seqüências de bases (Adenina, Guanina, Citosina e Timina) e do chamado Dogma Central da Biologia Molecular, que explica como a célula transforma informações contidas no DNA em proteínas ou enzimas. O DNA contido em nossas células contém informações precisas sobre as seqüências de bases que geram cada proteína, e como e quando essas proteínas devem ser produzidas. Fazendo uma analogia, poderíamos ver o DNA como o software (englobando dados e programa), e a célula com o hardware biológico. Com o seqüenciamento dos genomas de várias espécies, temos as listas de bases que estão guardadas nos núcleos das células de indivíduos dessas espécies. Conhecendo o DNA (software) e o funcionamento da célula (hardware), poderíamos entender/prever o comportamento fisiológico dos seres vivos. Isso possibilitaria, por exemplo, o projeto de remédios e terapias muito mais eficazes. Porém, ainda há um grande caminho a ser percorrido. Este projeto tem como objetivo dar uma contribuição à essa área.

Biologia Sistêmica [Chong e Ray 2006, Ideker et al 2001] é o estudo de um organismo visto como uma rede integrada e interativa de genes, proteínas e reações químicas que dão origem à vida. Ao invés de analisar componentes ou aspectos individuais do organismo, como metabolismo de açúcar ou o núcleo celular, os biólogos desta área focam em todos os componentes e na interação entre eles, tudo como parte de um único sistema. Estas interações são responsáveis pela forma e funções de um organismo. O sistema imunológico, por exemplo, não é o resultado de um único gene ou mecanismo. Na verdade, as interações de numerosos genes, proteínas, mecanismos e o ambiente externo do organismo produzem respostas imunológicas para combater doenças e infecções.

À medida que cientistas desenvolvem ferramentas e tecnologias que nos permitem procurar pelas bases da atividade biológica – genes e proteínas –, aprendemos que estes componentes quase sempre funcionam em conjunto. Eles interagem uns com os outros e com outras moléculas de maneiras altamente estruturadas mas incrivelmente

complexas. A Biologia Sistêmica procura entender estas interações complexas, pois elas são a chave para entender a vida. A função individual e a interação coletiva dos genes, proteínas e outros componentes de um organismo são frequentemente caracterizadas em conjunto como uma rede de interação. Na verdade, entender esta interação do genoma de um organismo e as influências ambientais externas é crucial para desenvolver o entendimento sistêmico de um organismo. A compreensão das interações de diversas hierarquias de informação biológica – DNA, RNA, proteínas, complexos macromoleculares, redes de sinalização, células, órgãos, organismos, espécies dentro de seu contexto ambiental – promoverá novos conceitos e inovações práticas.

Uma questão fundamental é como construir modelos abstratos fidedignos para analisar e prever o comportamento de sistemas tão complexos. Várias características básicas de sistemas biológicos (compostos por um número enorme de componentes, reatividade, mobilidade dos componentes, comportamento não-determinístico, etc.) são também características encontradas com frequência crescente em sistemas computacionais. Assim, técnicas de modelagem adequadas para esses sistemas computacionais poderiam ser utilizadas para modelar sistemas biológicos. Isso parece ainda mais natural se considerarmos que sistemas biológicos são também *programados*: o DNA contém o código que governa todo o comportamento desses sistemas (sendo sistemas essencialmente reativos, o controle se dá escolhendo as reações que devem ser realizadas).

Na Computação, a área de software evoluiu de sistemas muito simples, implementações de funções simples, para sistemas bastante complexos, com milhares de linhas de código em linguagens de alto nível. Estamos acostumados a desenvolver software em fases, usando técnicas de refinamentos e modularização. Na biologia molecular, os processos ainda são descritos ou através de conjuntos de equações matemáticas ou através de métodos ad hoc. Existem idéias de como modularizar ou descrever refinamentos, mas as técnicas de engenharia de software, principalmente as baseadas em métodos formais, poderiam contribuir muito nessa área. Já existem muitos modelos formais para simular redes de interações biológicas. Os métodos mais difundidos são baseados em equações diferenciais e suas variações [de Jong 2002]. Porém, estes métodos procuram modelar as redes em níveis de abstração bastante baixos, o que dificulta a modelagem de sistemas muito complexos (somente partes muito pequenas podem ser descritas assim, pois o número de variáveis equações seria imenso de outra forma). Modelos híbridos, como XS-Systems [Antoniotti et. al. 2003], ou simbólicos, como baseados em redes de Petri [Reddy 1994, Goss e Peccound 1998] são mais flexíveis nesta característica, mas em geral não permitem descrições modulares. Modelos baseados em álgebras de processos, como [Regev et. al. 2004, Pinto et. al. 2007, Danos e Krivine 2003, 2004, Leccal et. Al 2004, Curti et. al. 2004], descrevem bem o comportamento interativo do sistema. Estes modelos são adequados para descrever e analisar as séries de interações que um sistema pode apresentar. Mas, como não existe uma representação explícita dos estados, a descrição de alguns tipos de sistemas e propriedades dos mesmos (por exemplo, propriedades sobre estados alcançáveis) é difícil. Além disso, as linguagens de modelagem formal usadas são normalmente bastante difíceis de entender para os especialistas da área biológica, o que pode ser um grande problema, já que eles detêm o conhecimento da área na qual a modelagem computacional está sendo utilizada, e portanto são eles que devem validar o modelo proposto.

Algumas características parecem ser fundamentais para um método adequado para a

modelagem de sistemas biológicos: **Reatividade:** Componentes de um sistema biológico reagem a estímulos de outros componentes ou do ambiente. Modelos computacionais para sistemas reativos são a base de modelos para sistemas biológicos; **Não-determinismo:** Normalmente, existe mais de um caminho para uma reação ser atingida, bem como existe a possibilidade da reação não ser realizada. Assim, um modelo computacional adequado deve levar em consideração as probabilidades de ocorrência de cada ação; **Compartimentalização:** Os agentes dos processos biológicos interagem quando estão fisicamente próximos, em uma mesma região da célula. Como a célula é dividida em compartimentos, precisa-se levar em consideração o local onde o agente está para descrever seu comportamento (os agentes podem se mover entre os compartimentos).

Além dessas características, outras, como economia de energia e falhas nos processos (processos biológicos são bastante suscetíveis a erros), também podem ter papel importante na modelagem. Para que se consiga construir modelos abstratos realmente fidefignos dos processos biológicos, alguns requisitos do método de modelagem são necessários: **Modularidade:** A técnica básica para permitir a descrição de sistemas complexos é a divisão do modelo em componentes. Cada componente deve ter sua interface bem definida, e técnicas de abstração e refinamento devem ser usadas durante o processo de construção do modelo. **User-friendliness:** Qualquer modelo formal deve ser validado para ser útil. Pode-se utilizar técnicas de verificação formal para auxiliar neste processo (provando-se, por exemplo, que o modelo tem determinadas propriedades). Porém, seria extremamente importante que os especialistas da área biológica pudessem entender e discutir o modelo, ao invés de apenas acreditar que ele é realmente a abstração da realidade que se está tentando representar. Neste sentido, linguagens formais adequadas, ou seja, que facilitem o entendimento do modelo a profissionais desta área de aplicação, são um requisito desejável.

A definição de uma ou um conjunto de técnicas para modelagem de sistemas biológicos que permita a descrição integrada das características listadas acima, atenda os requisitos e ainda permita a análise do modelo gerado (usando não somente simulação, mas também técnicas de prova de propriedades sobre os modelos) é um grande desafio para a área de Computação. Os impactos da pesquisa nesta área são imensos, indo desde o melhor entendimento do nosso corpo (entendendo como a natureza “computa”), até o desenvolvimento de drogas e terapias realmente eficazes no tratamento de doenças (somente entendendo realmente um processo somos capazes de interferir nele para gerar com segurança os resultados que queremos).

6. Modelagem de Sociedades Financeiras Artificiais: Tópicos em Economia Computacional

Reproduzir resultados reais relativos a previsões do mercado financeiro através de algoritmos, tem sido o sonho de muitos cientistas, como matemáticos, cientistas de computação e físicos. Embora essa idéia seja tentadora, não menos importante é observar o comportamento emergente de aspectos financeiros do mercado ou mesmo de sociedades, fazendo simulações computacionais de agentes sob diferentes estratégias. Embora este último não traga uma importância imediata para muitos seres ela pode trazer importantes resultados que imitem sociedades reais em importantes aspectos, como a distribuição de riqueza, medidas das desigualdades sociais (através do chamado coeficiente de Gini¹), comportamento de grupos que tendem a investir em relação a bem

1 (Por exemplo: http://pt.wikipedia.org/wiki/Coefficiente_de_Gini)

públicos [R. da Silva et al 2006] dentre outros, o que geralmente pode ser feito "inventando" sociedades financeiras artificiais, sob certos vínculos e regras em uma rede ou mesmo em um regime onde todos interagem com todos, conhecido na literatura como regime de campo médio.

Dentre as principais vertentes onde supomos que cientistas de computação possam dar importantes contribuições para a modelagem ou análise de sistemas econômicos com grande necessidade computacional, destacamos três linhas principais: (1) Análise de séries financeiras reais; Neste caso usam-se modelos de ajuste e previsão para modelar séries de retorno, preço e outros importantes indicativos do mercado. Nesta linha encontram-se mais os estatísticos concentrando-se principalmente na inferência de parâmetros de por exemplo, distribuições dos retornos, autocorrelações em séries temporais financeiras dentre outros bem como utilizar estas séries para realizar previsões utilizando modelos de regressão [P. A. Morettin, 2006]. Nesta linha os cientistas de computação podem e devem interferir por exemplo com métodos de "data mining" dentre outros bem como a introdução de novos algoritmos de previsão, que levem em conta aspectos gerais e locais na previsão etc.

(2) Modelo multi-agente em sistemas econômicos: Aqui os cientistas de computação já vem dando importantes contribuições em conjunto com físicos e matemáticos [H. Takahashi, T. Terano 2003, E. Samanidou 2007]. Trata-se de modelos onde os agentes seguem regras para comprar e vender ações e então se pode estudar a dinâmica financeira que emerge deste sistema complexo. Neste caso pode-se colocar uma série de adendos, tais como aprendizagem por reforço, diferentes regras para os agentes decidir a compra ou venda etc. Ressaltamos como um ponto importante a ser simulado, sistemas com portfólio variado [L.A.M. Garcia, R. da Silva 2007] ainda foi pouco explorado na literatura e merece uma atenção especial, pois ainda não foi considerado em aclamados (veja por exemplo [H. Takahashi, T. Terano 2003, E. Samanidou 2007]). Definitivamente a idéia fundamental para quem quer trabalhar com economia computacional é trazer economistas para que estas idéias comecem a ser incorporadas na rígida teoria econômica, tradicionalmente desenvolvida apenas por economistas.

(3) A teoria evolucionária dos jogos para descrever sociedades financeiras: Neste caso pode-se estar interessados em simular qual o ganho médio e as possíveis flutuações disso em uma sociedade heterôgenea (multi-estratégica) cujas negociações por comida, ou por recursos, sejam descritas por protocolos prescritos pela teoria clássica de jogos e que evoluem temporalmente com evolução de estratégias [R.da Silva, G.A. Kellermann 2007, G. A. Kellermann, R. da Silva 2007]. A idéia é observar diagramas bi e tridimensionais de parâmetros que indicam regiões onde uma estratégia seja dominante sobre as outras. Além disso, pode-se estudar a evolução das densidades de estratégias e até mesmo a sobrevivência de estratégias em redes ou simplesmente em regimes de campo médio (onde um jogador interage com todos os outros jogadores) escolhidos aleatoriamente para enfrentarem-se. Sob um grande número de repetições é conveniente e importante observar as possíveis distorções destas versões evolucionárias de jogos com relação aos equilíbrios de Nash existentes na versão clássica.

7. Conclusões

Em resposta aos desafios propostos pela Sociedade Brasileira de Computação, este artigo propõe cinco linhas de pesquisa relevantes no desenvolvimento de sistemas computacionais complexos. Estas linhas de pesquisa têm impactos potenciais científicos e sociais altos, pois estão diretamente relacionados à cognição humana, ao

desenvolvimento de sistemas para a web, à biologia computacional, à econofísica e modelos sócio-econômicos, à interação do homem com computadores e com a natureza. A construção de sistemas complexos artificiais, biologicamente inspirados ou não, constitui-se em um grande desafio à ciência da computação e às demais ciências, pois evidências científicas e tecnológicas apontam para o aumento da interação entre a computação e virtualmente todos os domínios do conhecimento humano. No entanto, para que a hipotética promessa de impacto tecnológico seja realizada, é necessário que pesquisadores e profissionais estejam preparados para modificar e adaptar suas práticas de trabalho. Isto corrobora a necessidade de novas práticas de pesquisa, exigindo equipes multi-disciplinares e com forte conhecimento científico fundamental. Este conhecimento é essencial, na nossa opinião, para permitir a construção de sistemas computacionais complexos.

8. Referências

- Antoniotti, M., Policriti, A., Ugel, N., and Mishra, B. (2003). "Model building and model checking for biochemical processes". *Cell Biochem Biophys*, 38(3):271–86.
- Berwick, R. and Niyogi, P. (1985) "Learning from Triggers". *Linguistic Inquiry*, 27(4): 605-622.
- Boldi, P., and S. Vigna (2004), "*The WebGraph Framework I: Compression Techniques*", Proc. of the 13th World Wide Web Conference, p. 595-601.
- Briscoe, T. (1999) "The Acquisition of Grammar in an Evolving Population of Language Agents". *Machine Intelligence 16: Elec. Trans. in Artificial Intelligence*.
- Chomsky, Noam (1965) *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, Noam (1995) *The Minimalist Program*. MIT Press,.
- Chong, L. and Ray, L.B. (2002). Whole-istic Biology. *Science*, 295(5560):1661
- Curti, M., Degano, P., Priami, C., and Baldari, C. T. (2004). Modelling biochemical pathways through enhanced π -calculus. *Theor. Comput. Sci.* 325(1):111–140.
- Danos, V. and Krivine, J. (2003). Formal molecular biology done in CCS-R. In *Bio-Concur'03, satellite workshop of CONCUR'03*, ENTCS.
- Danos, V. and Krivine, J. (2004). Reversible communicating systems. In *CONCUR 2004 - Concurrency Theory*, volume 3170 of *LNCS*, pages 292–307. Springer-Verlag.
- de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103.
- O'Donovan, R., Burke, M., Cahill, A., Van Genabith, J., Way, A. (2004). *Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II Treebank*. Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics, p. 319-326.
- Garcez, A. S. and Lamb, L.C. (2006). "A Connectionist Computational Model for Epistemic and Temporal Reasoning". *Neural Computation* 18(7):1711-1738. MIT Press.
- Garcia, L.A.M. and da Silva, R. (2007), "A new look about statistical modelling of financial artificial markets via multi-agent systems". in preparation
- Gaume, B. (2004). "Balades Aléatoires dans les Petits Mondes Lexicaux". In *I3 Information Interaction Intelligence*, CEPADUES éd. : 39-96.

- Goss, P.J.E. and Peccoud, J. (1998). "Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets". *PNAS*, 95(12):6750–6755.
- Gulli, A. and Signorini, A. (2005), "*The indexable web is more than 11.5 billion pages*", Proc. of the 14th World Wide Web Conference, p. 902-903.
- Halpern, J.Y., Harper, R., Immerman, N., Kolaitis, P., Vardi, M.Y., Vianu, V. (2001) "On the Unusual Effectiveness of Logic in Comp. Science". *Bull. Symb. Logic*. 7(2):213-236.
- Hinton, Geoffrey E. (2005). "What Kind of Graphical Model is the Brain?". Proc. of the Intl. Joint Conf. on Artificial Intelligence (IJCAI-05), pp. 1765-1775.
- Hoare, C.A.R. and Milner, A.J.R. (2004) (eds.) "*Grand Challenges in Computing Research*". ISBN 1-902505-62-X. British Computer Society.
- Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.*, 2:343–372.
- Jurafsky, D. and Martin, J.H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice-Hall.
- Kellerman, G.A. and da Silva, R. (2007) "*Statistical aspects of payoff over mean field regime in a heterogeneous population of the ultimatum players*", preprint, Elsevier.
- Kumar, R., Novak, J., Raghavan, P., Tomkins, A. (2003), "On the bursty evolution of blogspace", Proc. of the 12th World Wide Web Conference, p. 568—576.
- Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., Upfal, E. (2000) "*Stochastic Models for the Web Graph*". Proc. of the 41st FOCS, p. 57-65.
- Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A. (1999), "*Trawling the Web for Emerging Cyber Communities*". Proc. 8th World Wide Web Conference, p. 403-416.
- Lamb, Luis C., Borges, R.V., Garcez, Artur S. (2007). "A Computational Cognitive Model for Temporal Synchronisation and Learning." 22nd Conf. AAAI-07. AAAI Press.
- Lecca, P., Priami, C., Quaglia, P., Rossi, B., Laudanna, C., and Costantin, G. (2004) "A stochastic process algebra approach to simulation of autoreactive lymphocyte recruitment". *SIMULATION* 80:273–288.
- MacWhinney, B. (2000) "*The CHILDES project: Tools for analyzing talk*". Erlbaum.
- Manning, C. and Schütze, H. (1999). "*Foundations of Statistical Natural Language Processing*". MIT Press.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. (1953) "Equations of State calculations by fast computing machines". *J. Chem. Phys.* 21, 1087.
- Moretin, P.A. (2006) "*Um curso em séries temporais financeiras*", 17 SINAPE, ABE.
- Pennock, D.M., Flake, G.W., Lawrence, S., Glover, E.J., Giles, C.L. (2002), "Winners don't take all: Characterizing the competition for links on the web", *PNAS* 99(8): 5207-.
- Pinker, S. (1984) *Language Learnability and Language Development*. Harvard University Press.
- Pinto, M.C., Foss, L., Mombach, J.C.M., and Ribeiro, L. (2007). Modelling, property verification and behavioural equivalence of lactose operon regulation. *Computers in*

Biology and Medicine, 37:134–148.

Pollard, C. and Sag, I.A. (1989). *Information-Based Syntax and Semantics*. CSLI Pub.

Raghavan, S. and Garcia-Molina, H. (2003), *Representing Web Graphs*, Proc. of the 19th Conference on Data Engineering. p. 405-416

Reddy, V.N. (1994). *Modeling biochemical pathways: a discrete event systems approach*. MS thesis, University of Maryland.

Regev, A., Panina, E.M., Silverman, W., Cardelli, L. and Shapiro, E. (2004). “Bioambients: an abstraction for biological compartments”. *Theoretical Computer Science* 325(1):141–167.

Richardson, M. and Domingos, P. (2006): “Markov logic networks”. *Machine Learning* 62(1-2): 107-136.

Sakas, W. and Fodor, J. (2001). *The Structural Triggers Learner*. *Language Acquisition and Learnability*. S. Bertolo ed., p. 172-233, Cambridge University Press,.

Samanidou, E., Zschischang, E., Stauffer, D., and Lux, T. (2007) “Agent-based models of financial markets”. *Rep. Prog. Phys.* 70, 409--450

da Silva, R., Bazzan, A.L.C., Baraviera, A.T., Dahmen, S.R. (2006) “Emerging collective behavior and local properties of financial dynamics in a public investment game”, *Physica. A*, 371, 610-626

da Silva, R., Bazzan, A.L.C., Baraviera, A.T., Dahmen, S.R. (2006), “Dynamics of a Public Investment Game: from Nearest-Neighbor Lattices to Small-World Networks”. *Lecture Notes in Economics and Mathematical Systems*, 584:211-233

da Silva, R. and Kellerman, G. A. (2007), “Analyzing the payoff of a heterogeneous population in the ultimatum game”. submitted to *Physical Review E*.

Stanley, H.E. (1971), *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press.

Takahashi, H., Terano, T. (2003), “Agent-Based Approach to Investors' Behavior and Asset Price Fluctuation in Financial Markets”. *JASSS*, 6(3).

Thagard, P. (1996). *Mind: Introduction to Cognitive Science*. MIT Press.

Turing, A.M. (1950). “Computing machinery and intelligence”. *Mind*, 59, 433-460.

Valiant, L.G. (2003). “Three problems in computer science”. *J. ACM*, 50(1):96-99.

Villavicencio, A. (2002) “*The Acquisition of a Unification-Based Generalised Categorical Grammar*”. PhD Thesis. Computer Laboratory, University of Cambridge,.

Wigner, E.P. (1960) “On the Unreasonable Effectiveness of Mathematics in the Natural Sciences.” *Communications in Pure and Applied Mathematics* 13:1-14.

Zambiasi, M., Presa, T.A., Buriol, L.S., Orengo, V.M. (2007) “*Análise quantitativa e temporal do Wikigrafo-PT*”, Anais da Escola Regional de Banco de Dados. p. 22-31.

Zhang, Y., Kordoni, V., Villavicencio, A., Idiart, M. (2006). “*Automated multiword expression prediction for grammar engineering*”. Proc. of the Workshop on Multiword Expressions:, p. 36–44, Sydney, Australia. Association for Computational Linguistics.