

NanoTrack – Sistema de Gerenciamento de Dados de Nanoestruturas com Plugins Inteligentes

André S. F. Silva, Paulo H. S. Batista, Artur M. M. Costa,
Alessandra C. Faria-Campos, Sérgio V. A. Campos and Omar P. Vilela Neto

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
(UFMG) Belo Horizonte – MG – Brazil

{andresfs, paulo.batista, artur.costa, scampos, alessa, omar}@dcc.ufmg.br

Abstract. *This work describes a tool to support nanoscience and nanotechnology laboratories based on the concept of LIMS (Laboratory Information Management Systems). The tool consists of a workflow capable of storing and managing data related to the synthesis of nanostructures, resources for data sharing and intelligent systems to infer parameters to be used in future synthesis of new nanostructures.*

Resumo. *Este trabalho descreve uma ferramenta para apoio a laboratórios de nanociência e nanotecnologia baseada no conceito de LIMS (Laboratory Information Management Systems). A ferramenta consiste de um workflow capaz de armazenar e gerenciar os dados relacionados à síntese de nanoestruturas além de recursos para o compartilhamento de dados e sistemas inteligentes para inferir parâmetros a serem utilizados em sínteses futuras de nanoestruturas.*

1. Introdução

O avanço tecnológico observado nas últimas décadas tem propiciado a geração, o armazenamento e a divulgação de uma grande quantidade de dados. Diversas áreas de pesquisa e desenvolvimento têm buscado alternativas eficientes para lidar com esta grande onda de informações. Entre estas áreas estão a nanociência e a nanotecnologia, que tratam do estudo e síntese de estruturas, dispositivos e sistemas controlando a forma e o tamanho na escala atômica e molecular. Recentemente, o interesse em nanociência e nanotecnologia tem aumentado devido ao grande potencial que esta área tem em propiciar benefícios em diversos campos. Entretanto, a nanotecnologia ainda carece de ferramentas computacionais que apoiem o pesquisador em suas atividades [de la Iglesia et al. 2011]. O objetivo principal deste trabalho é construir um sistema computacional para gerenciamento de dados de experimentos em nanotecnologia com uma plataforma de apoio ao pesquisador que possua funcionalidades adequadas à captura e análise de dados desta natureza, como atividades de síntese, ambiente de simulação e compartilhamento de informação. O sistema apresentado é uma adaptação do sistema proposto em [Melo et al. 2010]. O sistema proposto é baseado em um LIMS (*Laboratory Information Management System*), um sistema desenvolvido com o objetivo de facilitar o gerenciamento e armazenamento de dados de laboratório que contempla todo o ciclo de vida dos dados e permite o controle do acesso dos usuários aos resultados armazenados [Hinton 1995]. Além disso, o ambiente de simulação e descoberta de conhecimento proposto é constituído por sistemas de inferências de dados, tais como Redes Neurais Artificiais (RNA), que são encapsulados em módulos complementares e incorporados à estrutura do LIMS como *plugins*. O

encapsulamento de cada simulador como *plugin* permitirá a evolução de cada sistema de inferência independente da evolução do LIMS, além de permitir que outros simuladores sejam integrados a ferramenta posteriormente. O novo sistema, denominado *NanoTrack* que foi prototipado em [Silva et al. 2012], supre as principais necessidades de gerenciamento de dados de laboratórios de nanotecnologia, provendo uma ferramenta eficiente e segura para a utilização por pesquisadores da área. Nesta nova versão, o *NanoTrack* foi acrescido de *plugins* que permitirão a incorporação de algoritmos de inteligência computacional que irão compor o ambiente de simulação proposto.

2. Nanotrack

Visando suprir a demanda de sistemas computacionais capazes de atender os laboratórios de nanotecnologia, neste trabalho foi desenvolvida a ferramenta *NanoTrack*, uma ferramenta orientada a workflows que utiliza como base o sistema Flux®[Melo et al. 2010]. Este sistema é um LIMS desenvolvido em Java com SGBD MySQL e interface *WEB* para acesso amigável. A ferramenta apresenta sistemas inteligentes de inferência e previsão integrados ao *workflow* como *plugins* que permitirão a aplicação de técnicas específicas de inferência quando necessário para auxiliar os especialistas na síntese de futuras nanoestruturas.

2.1. Modelagem

Cada laboratório possui características próprias de funcionamento e estas devem ser representadas em um modelo de dados a ser lido pela ferramenta. O primeiro passo para realizar a modelagem do laboratório é um entendimento da rotina de atividades do mesmo, com todas as suas etapas. A rotina do laboratório é então modelada em forma de um fluxo de trabalho (*workflow*), onde as etapas identificadas dentro dos processos existentes nos laboratórios são representadas como atividades. A seguir, toda a informação obtida durante as interações da especificação de requisitos é codificado no formato XPDL (*XML Process Definition Language*) [Interface 2002] para ser interpretada pelo *NanoTrack*.

2.2. Plugins

Para integração dos sistemas de inferência ao *NanoTrack* foram utilizados *plugins*. O *plugin* é um módulo de extensão capaz de adicionar funções a uma ferramenta hospedeira. Os *plugins* são acoplados ao sistema a fim de prover funcionalidades específicas que não eram atendidas pela ferramenta, tornando o desenvolvimento destes componentes independente do desenvolvimento da aplicação.

2.3. Interação: Plugin x NanoTrack

Para que um *plugin* seja introduzido na aplicação deve ser criada uma nova atividade no *workflow*, contendo as informações que serão passadas para o processamento do *plugin*. Uma vez que os atributos das atividades são preenchidos e o procedimento é confirmado pelo usuário, a aplicação cria um processo para a execução do *plugin*. O funcionamento do *plugin* não importa para o *NanoTrack*. Para a aplicação ele funciona como uma caixa preta, onde os dados de entrada são inseridos e é esperada uma saída. A saída obtida durante a execução do *plugin* é armazenada em um arquivo texto, que foi especificado como parâmetro pelo *NanoTrack*. Este arquivo texto é posteriormente utilizado para exibir o resultados encontrados para o usuário da aplicação.

3. Estudo de Casos

Neste projeto foram desenvolvidos três estudos de casos de laboratórios de nanotecnologia: **Pontos Quânticos, LabSem - Laboratório de Semicondutores (PUC-Rio)** e **Nanotubos de Carbono (Física-UFMG)**, sendo que, por limitação de espaço, apenas o primeiro será descrito neste trabalho.

3.1. Workflow do Laboratório de Síntese de Pontos Quânticos

A modelagem deste *workflow*, utilizado como protótipo, foi baseada na experiência anterior em utilização de dados de Pontos Quânticos [Singulani et al. 2008], onde as etapas do processo foram identificadas e organizadas de forma ordenada. Estas etapas podem ser classificadas em dois grupos: síntese (Síntese) e caracterização (Raio-X e Microscopia AFM). Além destes, um terceiro grupo de atividades foi inserido formando a parte dos módulos incorporados à ferramenta, os *plugins* (Consulta Externa, Rede Neural Treinada e Rede Neural Treinamento). Para cada processo identificado foi criada uma atividade que o representasse. Estas atividades são compostas por propriedades que representam os dados gerados durante a sua execução e estas deverão ser armazenados no banco de dados da ferramenta a fim de possibilitar futuras consultas e auxiliar os algoritmos de inferência na descoberta de conhecimento. Estas propriedades foram obtidas por meio do trabalho [Singulani et al. 2008], através de uma planilha eletrônica contendo um cadastro de diversos experimentos sobre a síntese de pontos quânticos. Após modelar e identificar as propriedades do processo de experimentação de Pontos Quânticos é preciso codificá-lo para o formato XPD L que será interpretado pelo *NanoTrack*. Para isto a ferramenta *Together Workflow Editor (TWE)* foi utilizada. Durante a criação das atividades no TWE, as propriedades identificadas anteriormente são codificadas em forma de atributos. A atividade “Consulta Externa” implementa um *plugin* que faz a busca por amostras cadastradas no *Nanotrack* que possuam os valores das propriedades desejadas, considerando um desvio. As amostras dentro do intervalo definido são apresentadas ao usuário de forma ordenada, priorizando aquelas mais próximas do valor desejado. Neste *workflow* o ambiente de simulação e descoberta de conhecimento é composto por dois *plugins*, um para *treinamento da RNA* e outro que é a *RNA treinada*, ambos implementados em *Matlab*. Através da utilização do simulador, o usuário da ferramenta poderá ajustar os parâmetros de entrada à sua maneira e o simulador irá processar estes dados e retornar o resultado.

3.2. Plugin de Treinamento da Rede Neural Artificial

Este *plugin* é utilizado para treinar uma RNA do tipo *Perceptron* de Múltipla Camadas. As amostras que serão utilizadas no treinamento são previamente escolhidas durante o seu cadastro na ferramenta e, com base nesta informação, o *plugin* irá gerar um arquivo que servirá de suporte para o treinamento da RNA. Neste estudo de caso, três redes neurais foram treinadas para inferir a **densidade**, a **altura média** e o **desvio padrão** do processo de crescimento de pontos quânticos. A rede obteve resultados com altíssima precisão, sendo que o erro percentual absoluto médio durante a etapa de teste foi de aproximadamente 2% e o coeficiente de correlação é próximo de 1, o que indica que o simulador é capaz de gerar resultados muito próximo do mundo real. Os resultados da rede neural desenvolvida para a inferência de **altura média** foram satisfatórios com um erro médio de aproximadamente 11% durante a etapa de teste e correlação de 0,97. Já os resultados

da rede neural para inferência de **desvio padrão** não foram satisfatórios, apresentando um erro médio de quase 30% e uma correlação de 0,15. A RNA para inferência de desvio padrão não foi aprimorada devido a falta de precisão das amostras cadastradas, entretanto o objetivo maior era mostrar a capacidade de integração do ambiente de simulação com o LIMS.

3.3. Plugin de Uso da Rede Neural Artificial

Nesta etapa, a melhor rede neural obtida no passo anterior, que será utilizada como simulador, é carregada na memória do *Matlab*. Com isto, todo o ambiente que foi gerado durante a criação da rede neural fica disponível. Para utilizar a rede neural o usuário deverá informar os valores de entrada que ele deseja simular. A atividade então invoca o *plugin*. Como os dados utilizados durante o treinamento foram normalizados, é importante que os dados a serem simulados também sejam normalizados. Após a normalização dos dados, a Rede Neural é invocada e os valores de entrada, são passados por parâmetro. A inferência é realizada de maneira extremamente rápida e o resultado obtido pela rede neural fica disponível, bastando apenas realizar o processo inverso da normalização para que os valores obtidos sejam condizentes com os valores reais.

4. Conclusão

Com o objetivo de suprir as carências aqui apresentadas, neste trabalho foi desenvolvido o *NanoTrack*, um LIMS, baseado na plataforma Flux®), responsável por gerenciar os experimentos realizados em um laboratório de nanotecnologia, a fim de apoiar a criação de nanomateriais. Também foi mostrado que o *NanoTrack* pode incorporar novas funcionalidades através da utilização de *plugins*, e através deles foi criado um ambiente de simulação. O desenvolvimento do sistema *Nanotrack* apresenta uma solução que poderá contribuir significativamente para a pesquisa em nanotecnologia ao auxiliar na organização e análise dos dados obtidos e propor sistemas de inferência automatizados que facilitarão a seleção dos dados para análise.

Referências

- de la Iglesia, D., Harper, S., Hoover, M., Klaessig, F., Lippell, P., Maddux, B., Morse, J., Nel, A., Rajan, K., Reznik-Zellen, R., et al. (2011). Nanoinformatics 2020 roadmap.
- Hinton, M. (1995). Laboratory management systems. *New York*.
- Interface, P. (2002). Xml process definition language. *Document Number WFMC-TC-1025 Document Status-XPDL*, 1.
- Melo, A., Faria-Campos, A., DeLaat, D. M., Keller, R., Abreu, V., and Campos, S. (2010). Sigla: an adaptable lims for multiple laboratories. *BMC genomics*, 11(Suppl 5):S8.
- Silva, A. S., Batista, P. H., Andrade, T. K., Vilela Neto, O. P., Campos, S., and Faria-Campos, A. (2012). Nanotrack – sistema inteligente de gerenciamento de dados de síntese de nanoestruturas. *Congresso Brasileiro de Ciência da Computação*.
- Singulani, A., Vilela Neto, O., Aurélio Pacheco, M., Vellasco, M., Pires, M., and Souza, P. (2008). Computational intelligence applied to the growth of quantum dots. *Journal of Crystal Growth*, 310(23):5063–5065.