

Curriculum Vitae

Personal Data

Name: Thiago Cunha de Moura Salles

Birthday: 10/30/1984

Gender: Male

Age: 26

Address: Av. Itaituba, 363, apt 101, Boa Vista

Zip Code: 31060-420

City/Country: Belo Horizonte / Brazil

Email: tsalles@dcc.ufmg.br

Skills

- Programming:**
- Fluency on C/C++.
 - Good level in Java, SQL and scripting languages (BASH, awk, perl and php).
 - Knowledge in the following tools/libraries: Valgrind, Google's Perf Tools and Protocol Buffers, OpenMP, cURL, Boost and HadoopPipes.
 - Knowledge in some noSQL databases (as the key-value based Tokyo Cabinet and BerkeleyDB, and the tabular-based Hypertable - BigTable's clone).
 - 1 year experience with design and implementation of distributed and parallel algorithms for Data Mining and Information Retrieval tasks.

Versioning System: SVN and GIT

- Natural Language:**
- Portuguese (mother tongue)
 - English (speak, read, write)

Typography: LaTeX mastered

UNIX Experience: Currently the administrator of a medium-sized Linux-centered academic laboratory (Laboratory of the UFMG Database Group, LBD). Some administrative tasks on a medium-sized cluster of the Laboratory for Treating Information (LATIN).

Scholarship

- **Master Degree**

Institution: Universidade Federal de Minas Gerais

Course: Computer Science

Dissertation: Automatic Document Classification Temporally Robust (Machine Learning)

Advisor: Prof. Dr. Marcos André Gonçalves

Location: Belo Horizonte, Brazil

Begin: Mar/2009

End: Mar/2011 (ongoing)

- **Bachelor Degree**

Institution: Universidade Federal de Minas Gerais

Course: Computer Science

Monograph: A SQL-based System to Support Timetabling (Databases)

Advisor: Prof. Dr. José Monteiro da Mata

Location: Belo Horizonte, Brazil

Begin: Mar/2004

End: Mar/2008

- **English Course**

Institution: Inglês & Cia.

Course: English (Advanced)

Location: Belo Horizonte, Brazil

Begin: 1998

End: 2002

- **Technical Course**

Institution: SENAI-MG

Course: Servicing, Assembly, Troubleshooting and Repair for microcomputers

Location: Belo Horizonte, Brazil

Begin: Feb/2000

End: Mar/2000

Professional Experience

Institution: Synergia/UFMG

Role: Java Developer (implementation of use cases and bug fixing)

Activities: Participated in several mini-courses (Spring, Hibernate and Struts frameworks, Unit Testing, UML 2.0) and in the development of the "Portal de Compras de Minas Gerais", a portal that allows public consultation of the purchases made by the Minas Gerais' Government.

Date: 2008 to 2009

Research Experience

- **Research Interests:**

Machine Learning, Information Retrieval, Databases

- **Publications:**

SALLES, T.; ROCHA, L.; PAPPA, G. L.; MOURÃO, F.; GONCALVES, M. A.; MEIRA JR, W. Temporally-Aware Algorithms for Document Classification. In the 33rd Annual International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR). Geneva, 2010.

SALLES, T., ROCHA, L., MOURÃO, F., PAPPA, G., CUNHA, L., GOMCALVES, M. A., MEIRA JR, W. Automatic Document Classification Temporally Robust. JIDM 1(2): 199-212 (2010).

PALOTTI, J., **SALLES, T.**, PAPPA, G., ARCANJO, F., GONCALVES, M. A., MEIRA JR, W. Estimating the Credibility of Examples in Automatic Document Classification. JIDM 1(3): 439-454 (2010)

SALLES, T.; ROCHA, L.; MOURÃO, F.; PAPPA, G. L.; CUNHA, L.; GONCALVES, M. A.; MEIRA JR, W. Classificação Automática de Documentos Robusta Temporalmente. In the XXIV Simpósio Brasileiro de Banco de Dados (SBBDD). Fortaleza, 2009.

PAPPA, G. L.; ZADROZNY, B.; ROCHA, L.; **SALLES, T.**; MEIRA JR, W.; GONCALVES, M. A. Exploiting Contexts to Deal with Uncertainty in Classification. In the First ACM SIGKDD Workshop on Knowledge Discovery from Uncertain Data (UKDD). Paris, 2009

MOTA, H.; ROCHA, L.; **SALLES, T.**; VASCONCELOS, F., H. Partial discharge signal denoising with spatially adaptive wavelet thresholding and support vector machines. Electric Power Systems Research, 2011

- **Event Participations:**

- XXXIII International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). 2010.
- XXV Simpósio Brasileiro de Bancos de Dados (SBBD). 2010.
- IX Simpósio de Fatores Humanos em Sistemas Computacionais (IHC). 2010.
- VII Simpósio Brasileiro de Sistemas Colaborativos (SBSC). 2010.
- XVI Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA). 2010.
- XXIV Simpósio Brasileiro de Banco de Dados (SBBD). 2009.

- **Grants and Awards:**

- M.Sc. Student Researcher Scholarship from CNPq (National Science and Technology Foundation - Brazil), 2009 to present.
- Donald B. Crouch Travel Grant and SIGIR 2010 Student & New Researcher Travel Grant, 2010.
- Second Best paper in SBBD, 2009.
- Second Best paper in SBBD, 2010.
- Best ongoing M.Sc. work at the "Semana de Seminários do PPGCC/DCC", 2010.

- **Project Participations:**

- *INFOWEB: Construction of a web search engine. Now a sub-project of the INWEB project, as described below (item 2).*
- *INWEB (National Institute of Science and Technology for the Web)*

1. PBCT: Brazilian Portal of Science and Technology. Aims at studying collaboration networks on science and technology involving Brazilian researchers that participate in the National Institutes of Science and Technology Program. Currently, responsible for the web crawler module and HTML extractors (parsing and extraction of the desired information units for further processing).

Modules written in C++, with cURL, Boost, OpenMP and ProtocolBuffers libraries.

Coordinator: Prof. Dr. Alberto Henrique Frade Laender

2. Brazilian Search Engine: This sub-project, also part of the INWEB aims at fetching, processing and providing documents from the Brazilian web, by means of a Search Engine. Our main goal is to develop a competitive search engine with state of the art technologies. Currently, responsible for the webpage fetching module (which addresses network errors, respects the Robots Exclusion Protocol, keeps track of useful information for the other modules, among others) and URL rewrite rules generation module (to avoid fetching already crawled (near-)duplicate pages).

Modules written in C++, with cURL, Boost, OpenMP and HadoopPipes libraries.

Coordinator: Prof. Dr. Nivio Ziviani

References

References are available upon request.