andré**santos**
nlp, webdev and sysadmin

about

mail
andrefs@andrefs.pt
homepage
http://andrefs.pt
twitter
@about_andrefs
github
andrefs
cpan
andrefs

languages

native portuguese
fluent english
spanish & french
notions

programming

♥ Perl
(Moose, Dancer)
Bash, LaTeX

interests

open source software, Perl, natural language processing, information retrieval, bioinformatics

education

since 2010   **MSc in Bioinformatics**   University of Minho, Portugal
Mining biological parameters from literature: an application to environmental decision support systems

2009–2011   **MSc in Natural Language Processing**   University of Minho, Portugal
Contributions for building a Corpora-Flow System
Development of a system to support automatic handling of corpora (cleaning, validation, format conversion), document alignment and calculation alignability metrics of several types of documents.

2006–2009   **BSc in Software Engineering**   University of Minho, Portugal
180 ECTS, final grade 15/20

experience

since 2011   **CEHUM, University of Minho**   Researcher
Project Per-Fide: Portuguese in parallel with six languages:
Español, Russian, Français, Italiano, Deutsch, English

since 2011   **GroubBuddies**   System administrator
Startup company which develops web-based solutions for groups and organizations. Typical tasks include:

- deployment and management of several Unix-based servers, with Apache and MySQL services
- managing Ruby on Rails applications on a production environment
- managing Git repositories
- development of helper scripts to automate and facilitate server management

2009 Jul-Sep   **Sapo Summberbits'09**   Research Internship
Open source software development
Development of natural language processing tools to help in general orthography migration challenges, motivated by the 1990 Portuguese Language Orthographic Agreement.

technical skills

Perl   Proficient
Including Moose, Catalyst, Dancer, DBIx::Class and other Modern::Perl concepts.

Web development   Proficient
Mostly using Perl frameworks, but also Bootstrap from Twitter, CSS, Javascript, CSS.

Ontologies and structured documents   Experienced
XML, XQuery, XPath, JSON and ontology formats: OWL, OBO, RDF.

	Systems administration, programming tools and methods	Experienced
	Bash, Subversion, Git, LaTeX, Apache, Agile.	
	Other programming languages	
	SQL, C, Java, PHP, UML, Haskell, Ruby, Pascal.	

training

2011 Jul	**Band and Olufsen Summer School**	Struer, Denmark
	Conceptual Design and Development of Innovative Products	
2010/11	**IdeaLab -- Business Ideas Lab**	TecMinho
2010 Jul	**Luso-Brazilian Evolutionary Computing School**	University of Minho, Guimarães
2010 Jun	**Catalyst 5.80**	Portuguese Perl Workshop'10
2005 Set	**Physics Summer School**	Faculty of Science, University of Porto
2005 Jul	**German Open Course**	Faculty of Arts, Univerity of Porto

other activities

CeSIUM University of Minho's Software Engineering Student Center

2009/10	**Vice-president**	http://cesium.di.uminho.pt
2010/11	**Director at CAOS**	http://caos.di.uminho.pt
	CeSIUM's Open Source Support Center	

Programming Challenges

2012	**APPP Perl Code Sprint**	http://perl.pt
	3rd qualified in the Perl programming competition organized by the Portuguese Perl Mongers group.	
2009	**Microsoft Imagine Cup'09**	Lisbon, Portugal
	5th classified in the national finals.	
2008	**ACM South Western European Regional Contests**	Nuremberg, Germany
	International programming competition.	
2007-2009	**Inter-University Programming Marathon**	IST'07, UC'08, ESTGA'09
	National ACM-style programming competition.	

Sports

since 1999	**Judo**	INATEL; University of Minho
2010-2012	**National Judo University Championships**	Aveiro'10, Lisbon'11, Aveiro '12
2007/08	**Basketball**	University of Minho
1998-2000	**Artistic gymnastics**	INATEL

publications

articles in peer-reviewed journal

Applying a text mining framework to the extraction of numerical parameters from scientific literature in the biotechnology domain
A. Santos, R. Nogueira, A. Lourenço
Advances in Distributed Computing and Artificial Intelligence Journal (ADCAIJ) *1 (2012) pp. 1–8. 2012*

peer-reviewed conferences/proceedings

Structural alignment of plain text books
André Santos, José João Almeida, Nuno Carvalho
Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 2012, Istanbul, Turkey

A survey on parallel corpora alignment

A. Santos

MI-Star 2011, Braga, Portugal, 2011

Text::Perfide::Bookcleaner, a Perl module to clean and normalize plain text books

A. Santos, J.J. Almeida

Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN2011), Huelva, Spain, 2011

Bigorna – A Toolkit for Orthography Migration Challenges

J.J. Almeida, A. Santos, A. Simoes

Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010, Valletta, Malta

AGile, a structured editor, analyzer, metric evaluator, and transformer for Attribute Grammars

A. Rocha, A. Santos, D. Rocha, H. Silva, J. Mendes, J. Freitas, M. Coelho, M. Regedor, D. Cruz, P.R. Henriques

INForum 2010, Braga, Portugal, 2010

dissertations

Contributions for building a Corpora-Flow system

A. Santos

Master's dissertation, Departamento de Informática, Escola de Engenharia, Universidade do Minho, 2011

presentations

Cleaning plain text books with Text::Perfide::BookCleaner

A. Santos

In Portuguese Perl Workshop'11, Lisbon, September 2011

Preparação de livros para alinhamento

A. Santos

In I Jornadas Internacionais Per-Fide: Corpora e Tradução, 2011

Detecção e Correcção Parcial de problemas na Conversão de Formatos

A. Santos

In I Workshop Per-Fide: Construção, exploração e aplicação de Corpora Paralelos. 2010

Mini aplicações web com Mojolicious::Lite

A. Santos

In Portuguese Perl Workshop'10, Porto, June 2010