

# *Globs*: Gestor Cooperativo de Glossários e Terminologia

Aluno: Nuno Veloso <nunoveloso18@gmail.com>  
Orientador: Alberto Simões <ambs@di.uminho.pt>

Julho 2008

## 1 Orientador

### Nome:

Alberto Manuel Brandão Simões <ambs@di.uminho.pt>

### Instituição:

Departamento de Informática (<http://www.di.uminho.pt/>)  
Universidade do Minho (<http://www.uminho.pt/>)

### Experiência:

Alberto Simões é professor convidado no Departamento de Informática da Universidade do Minho onde realizou o seu Mestrado e Doutoramento na área de Processamento de Linguagem Natural. Lecciona neste departamento há mais de 8 anos, tendo vindo a colaborar em diversos projectos de onde se salienta:

- dicionário e analisador morfológico jSpell;
- dicionários da língua portuguesa para várias ferramentas open-source, como o Firefox, Thunderbird, ispell, aspell e hunspell;
- desenvolvimento da aplicação NumExp (<http://www.numexp.org/>);
- desenvolvimento de vários módulos Perl disponíveis no CPAN (<http://search.cpan.org/~ambs>).

Mais informação na sua página pessoal: <http://ambs.perl-hackers.net/>.

**Proposta:** <http://natura.di.uminho.pt/wiki/doku.php?id=ferramentas:globs>

## 2 Aluno

### Nome:

Nuno André Veloso Matias <nunoveloso18@gmail.com>

### Instituição:

Licenciatura em Engenharia Informática (<http://lei.di.uminho.pt/>)  
Departamento de Informática (<http://www.di.uminho.pt/>)  
Universidade do Minho (<http://www.uminho.pt/>)

### Curriculum Resumido:

Estudos primários e secundário em Bruxelas, Bélgica. Obtenção do 12<sup>o</sup> na área das Humanidades, com destaque para a Matemática, as Ciências (Biologia, Física e Química) e realização de um Trabalho de Final de Estudos defendido perante um júri. Formação enciclopédica e aberta ao multiculturalismo. Frequência do 1<sup>o</sup> ano da Licenciatura em Engenharia de Redes e Sistemas Informáticos (Faculdade de Ciências da Universidade do Porto). Frequência dos 1<sup>o</sup>, 2<sup>o</sup> e 3<sup>o</sup> anos das LESI/LEI na Universidade do Minho. Mais informações em <http://www.nunoveloso.org>.

### Experiência:

Desde 2005, membro da equipa técnica (voluntário) do Centro de Apoio ao *Open Source* (<http://caos.di.uminho.pt>), UMinho. Principais actividades e responsabilidades: assegurar manutenção do sistema informático do Centro de Estudante, assim como auxiliar a comunidade *Open Source*. Partilhar conhecimentos através eventos pedagógicos como instalações de Linux (assistente técnico) e um *workshop* de L<sup>A</sup>T<sub>E</sub>X (orador).

Desde 2007, coordenador principal (voluntário) e fundador da *Apple Tech Team* (<http://www.appletechteam.org>), UMinho. Principais actividades e responsabilidades: coordenador principal e fundador, relações públicas, coordenador de eventos, *webmaster*, administrador de sistema.

Desenvolvimento de código no âmbito do curso em linguagens como Haskell, C, Prolog, Java, PHP; conhecimentos de Pascal e QBasic (1999/2000). Todo o código desenvolvido é *open source* e documentado. Envolvimento com a comunidade de *developers* de *Drupal*.

### Métodos de Desenvolvimento:

Análise e especificação dos requisitos (*brainstorming*, dados e pedidos, agregação de conceitos necessários à realização);

Concepção e desenho da resolução (estruturas de dados, diagramas de apoio [ERs de BD, sequência, fluxo, actividades, etc] e injeção de conceitos necessários à resolução);

Codificação e testes (codificação estruturada e documentada — *Doxygen*, código fonte genérico servindo como API para implementações futuras, testes e *debugging* adaptados e verificação de código).

**Motivação:**

O objectivo deste projecto reside na implementação de uma aplicação Web que respeite todas as características de uma ferramenta electrónica de criação e gestão de terminologia. Para além desses requisitos, pretende-se ainda que o gestor siga o conceito de Web 2.0 (por exemplo edição colaborativa), assim como permitir exportar o conteúdo.

O interesse portado pelo aluno sobre as línguas e culturas estrangeiras assim como o contacto prévio com aplicações de gestão de glossários revelaram-se um facto decisivo na escolha do projecto. A utilização do conceito de Web 2.0 e o desenvolvimento Web usando uma tecnologia que não *.net* – demasiado usada no mercado – também pesou fortemente na escolha.

## 3 Projecto

**Licença:** GNU General Public License, v.3;

**Seleção:** O aluno tem vindo a desenvolver a aplicação em causa nos últimos dois meses numa disciplina da Licenciatura em Engenharia Informática, da Universidade do Minho, tendo obtido resultados bastante motivadores;

**Calendarização** A calendarização do projecto, objectivos propostos, contextualização e descrição detalhada do projecto serão apresentados nas próximas secções;

**Interacção:** De acordo com a descrição do projecto (ver secções seguintes), o aluno terá de interagir com a comunidade Drupal, uma vez que todo o desenvolvimento da aplicação assenta nesta ferramenta. Além disso, será necessária a interacção com algumas instituições que definem standards abertos como a LISA: Localization Industry Standards Association.

**Motivação:** A principal motivação para o desenvolvimento deste projecto prende-se com a sua grande utilidade, não só para o público em geral, mas também para os alunos das diversas licenciaturas de tradução que poderão tirar partido desta ferramenta para a construção dos seus glossários. Em particular, na Universidade do Minho pretende-se usar esta ferramenta nas disciplinas de terminologia da Licenciatura em Línguas Aplicadas.

### 3.1 Contextualização

No terceiro ano da Licenciatura em Engenharia Informática, da Universidade do Minho, foi proposta a construção de uma ferramenta Web para a gestão cooperativa de terminologias e glossários. O aluno Nuno Veloso integrou um grupo que desenvolveu (a par das

restantes disciplinas) um protótipo básico deste sistema, e que está actualmente disponível em <http://eremita.di.uminho.pt/terminologia/>. Durante o desenvolvimento o Nuno destacou-se do resto do grupo, tendo sido o responsável pela maior parte do seu desenvolvimento. Deste modo, torna-se natural o convite para a continuação do desenvolvimento desta aplicação.

Esta aplicação surgiu face à necessidade de ensinar alunos de licenciaturas de línguas e tradução a construção de glossários. Actualmente não existem ferramentas de código-aberto (ou livres) com este objectivo, sendo que as ferramentas comerciais usadas na indústria são demasiado caras para poderem ser adquiridas por uma Universidade.

É também importante salientar que não se pretende o desenvolvimento de uma ferramenta apenas para a Universidade do Minho: não só existem contactos com outras Universidades que estão interessadas numa ferramenta deste género, como os próprios profissionais da área da tradução e línguas poderão tirar partido da ferramenta.

## 3.2 Globs

O *Globs* é uma aplicação *web* desenvolvida sobre o Framework *Drupal*, e pretende ser um gestor de terminologia/glossários cooperativo.

Na área da tradução, uma terminologia ou glossário é visto como um conjunto de registos relativos a determinado conceito. Para cada um destes conceitos existem instâncias para cada língua (vulgo identificador do conceito). Além desta informação, são associadas propriedades ao conceito e/ou a cada uma das instâncias.

Como exemplo ilustrativo, o conceito *gato* (animal doméstico peludo) pode ser representado por vários identificadores, como *gato* ou *bichano* para a língua portuguesa, *cat* para a língua inglesa, etc. Para cada uma destas línguas podem-se associar propriedades. Por exemplo, o género e o número da palavra, mas também o estilo (formal ou coloquial), exemplos de uso, etc.

O *Globs* pretende implementar um sistema para a gestão deste género de terminologias, em que a estrutura (também conhecida por *micro-estrutura*) pode ser definida de forma independente para cada terminologia de acordo com os objectivos para os quais está a ser desenvolvida.

Actualmente o *Globs* inclui as seguintes funcionalidades básicas:

- gestão de utilizadores;
- gestão de permissões sobre glossários;
- criação básica de glossários orientados ao conceito;

- adição de línguas e propriedades a cada conceito do glossário;
- comunicação inter-colaboradores de um glossário;

### 3.3 Objectivos

Embora o *Globs* actual funcione de forma básica, é imprescindível a adição de novas funcionalidades para que a ferramenta seja, realmente, útil:

- **Internacionalização/localização:**  
Toda a aplicação foi desenvolvida em Inglês, mas deverá ser tornada multi-lingue. A própria tarefa de internacionalização/localização pode tirar partido da própria aplicação;
- **Estruturação de Glossários:**  
Adicionar uma estrutura rígida (e independente) a cada glossário.
  - que línguas podem ser instanciadas;
  - que atributos/propriedades podem ser utilizadas;
  - que tipo de dados está associado a cada atributo/propriedades (nomeadamente, *pick-lists*);
  - definição de relações inter-conceito;
- **Intercâmbio de Glossários:**  
Adicionar funcionalidades de intercâmbio com outras aplicações.
  - exportação de vistas sobre o glossário em PDF;
  - exportação de vistas hierárquicas do glossário no formato ISO-2788/5964 e/ou no formato OWL;
  - importação e exportação do formato aberto TBX (Term base exchange);
  - importação e exportação do formato textual MultiTerm;

### 3.4 Calendarização

Os objectivos definidos na secção anterior são demasiado ambiciosos para os três meses do Summerbits08. Sugue-se uma calendarização dos objectivos mais prementes, dividindo-os pelas 12 semanas do programa:

- **Semana 1:** Colocação do código actual em <http://softwarelivre.sapo.pt>, criação da página do projecto e ambientação às ferramentas disponibilizadas pela SAPO;

- **Semana 2:** Sistema básico para a internacionalização da aplicação;
- **Semana 3:** Tradução da aplicação para a língua portuguesa. Permitir ao utilizador a escolha da língua a usar. Adaptação do esquema relacional da base de dados para suporte à configuração dos glossários;
- **Semana 4:** Implementação do novo esquema relacional, e adaptação do código actual.
- **Semana 5:** Adição da interface para definição da estrutura dos glossários. Adaptação da interface de edição de conceitos.
- **Semana 6:** Realização de testes.
- **Semana 7:** Criação de vistas (por língua) sobre o glossário em PDF, utilizando L<sup>A</sup>T<sub>E</sub>X como ferramenta de typesetting.
- **Semana 8:** Análise do formato textual usado pelo MultiTerm. Implementação de um parser.
- **Semana 9:** Implementação do importador de MultiTerm. Realização de testes.
- **Semana 10:** Análise do formato TBX;
- **Semana 11:** Exportador da estrutura do glossário para TBX;
- **Semana 12:** Exportador do glossário para TBX;

Durante todo este processo pretende-se manter o sistema a funcionar o máximo tempo possível uma vez que existem potenciais utilizadores que poderão servir de *beta-testers*.