



Refatoração do projeto Delpo

Adriano Tetsuaki Ogawa Santin
Luiz Fernando Antonelli Galati
Mauricio Luiz Abreu Cardoso

Supervisor: Prof. Dr. Marco Dimas Gubitoso

Bacharelado em Ciência da Computação



Introdução

Comparado com outras línguas românicas, o português é, de longe, o idioma que tem a informação etimológica mais incompleta e fragmentada.

Tendo isso em vista, foi criado, em 2012, o Núcleo de apoio à pesquisa em Etimologia e História da Língua Portuguesa (NEHiLP). Um de seus projetos é o Dicionário Etimológico da Língua Portuguesa (DELPO), cujo objetivo é satisfazer os requisitos da linguística moderna e sanar deficiências dos atuais léxicos etimológicos. O dicionário, já disponível online, funciona através de um complexo sistema de *software* e tem seus dados e todas as modificações em seus verbetes assinadas por pesquisadores.

Levando em consideração que os *softwares* atualmente responsáveis pelo funcionamento do projeto contêm problemas que são difíceis de serem tratados (como imperfeições na escrita de códigos, ausência de documentação que especifique as atribuições de cada parte do sistema, ferramentas antigas e ineficientes, etc.), propusemo-nos a desenvolver, neste trabalho, um novo DELPO que conte com um sistema computacional estável, consistente e bem documentado.

O que é exatamente o DELPO?

Como o próprio nome já indica, DELPO é um dicionário de etimologia.

Para explicar como ocorre o processo de inserção de palavras no dicionário, vamos utilizar a seguinte definição:

Dizemos que uma data a no formato $dd/mm/aaaa$ é menor do que outra data b no mesmo formato se o dia especificado por a tiver ocorrido antes do especificado por b .

Para adicionar palavras ao dicionário, é necessário que um pesquisador submeta uma obra, juntamente com sua data, a um programa chamado moedor. Cada palavra presente no texto inserido é analisada e, então, uma das seguintes ações é executada:

- se a palavra não estiver no banco de dados, o moedor nele a insere (juntamente com a data da obra submetida);
- se a palavra estiver no banco de dados e a data do texto submetido for menor do que a data da palavra já existente no banco, o programa atualiza a data da palavra do banco para a data do texto submetido.

Dessa maneira, ao final da inserção de uma ou mais obras, a cada palavra do banco de dados estará associada uma data que representará o dia exato da primeira ocorrência do vocábulo conhecida até então. Quanto mais obras forem submetidas ao moedor, maiores serão as chances de as datas associadas às palavras retratarem o dia exato em que cada uma delas ocorreu pela primeira vez em um texto da língua portuguesa.

Ferramentas utilizadas

Framework Ruby On Rails

O Rails conta com numerosas *libs*, conhecidas como *gems*, que ajudam em todo o processo de desenvolvimento de um *Web Service*. Há *gems* para a execução de tarefas como elaborar o mecanismo autenticação de usuários, utilizar o Bootstrap e realizar testes automatizados do backend e da interface.

A linguagem Ruby possui um grande leque de expressões regulares importantíssimo para a análise de palavras.

Banco de Dados PostgreSQL

É open-source e suporta grandes cargas de dados, o que é necessário em um projeto de longo prazo e que pretende abrigar muitas obras.

Trello

Permite a visualização de todo o projeto em um só lugar e facilita a comunicação e a divisão de tarefas entre todos os desenvolvedores.

Desenvolvimento

Essencialmente, o desenvolvimento do sistema realizou-se em três etapas.

Primeiramente, montou-se a estrutura da aplicação Rails e criou-se o Banco de Dados. Nesse estágio, foram geradas diversas tabelas capazes de atender às várias demandas do projeto, criou-se uma interface para manipulação do Banco e foram elaborados mecanismos para a autenticação dos usuários do sistema.

Em seguida, realizaram-se testes, tanto da aplicação quanto das tabelas. Nesse estágio, diversos erros e inconsistências foram encontrados e corrigidos.

Por último, escreveu-se o código do moedor. Nesse estágio, procedeu-se a um minucioso estudo do código legado (em linguagem Perl) do antigo DELPO. Muitas adaptações e renovações tiveram de ser realizadas para que a escrita no novo código (em linguagem Ruby) fosse possível.

Resultados

Framework robusto, com muitas libs e comunidade ativa
Código com alta escalabilidade e acesso a API's mais modernas
Estrutura MVC, dinâmica melhor para um serviço web