Projeto e implementação de serviços analíticos para o DataUSP-PósGrad

Bruno Padilha[1], João Eduardo Ferreira (Orientador)[2]

[1]IME-USP - Instituto de Matemática e Estatística da Universidade de São Paulo,
Departamento de Ciência da Computação

Introdução

O DataUSP-PósGrad é um conjunto de serviços analíticos RESTful, cujo propósito é auxiliar a Pró-Reitoria de Pós-Graduação da Universidade de São Paulo na tomada de decisões, por meio de relatórios *Ad hoc* com estatísticas diversas dos programas, áreas e pessoas vinculadas à pós-graduação.

Figura 1 : Tela inicial do sistema (uspdigital.usp.br/datausppg)

RESTful Web Services

Um serviço web RESTful oferece seus recursos na web por meio de URIs, que são os endereços HTTP que unicamente os identificam. Além disso o protocolo HTTP é usado não apenas como camada de transporte, mas também como protocolo de acesso ao serviço utilizando-se seus métodos básicos, como GET, POST, PUT, e DELETE. [4]

Figura 2 : Mapa de serviços agregados

Técnologia e arquitetura

O DataUSP-PósGrad tem duas grandes componentes: o servidor de recursos e a interface de interação com o usuário.

Servidor

O servidor de recursos foi construído em Java segundo a especificação JAX-RS, definida na *JSR*(Java Specification Request) 311, que nada mais é do que a especificação de *RESTful Web Services* em Java. A *API* utilizada foi o *Jersey* que é uma das implementações padrão da JAX-RS e permite o uso de anotações da linguagem. Criando assim uma aplicação altamente escalável em termos de vazão de dados e processamento e, ainda de fácil manutenção [5].
A comunicação com o servidor é feita majoritariamente utilizando-se mensagens *json*.

Interface

Já a interface de interação com o usuário foi construída com o que existe de mais moderno em tecnologia web, fazendo uso intensivo de HTML5 e JavaScript por meio do *JQuery* [6], além de chamadas assíncronas ao servidor de recursos com *Ajax* [7]. Esse conjunto de ferramentas computacionais possibilita a criação de gráficos tridimensionais animados em tempo real. Com esses gráficos, os usuários podem fazer análises de forma ágil e intuitiva"

Figura 3 : Exemplo de tela com relatório do tempo médio de titulação dos programas do IME

Mundo dos dados

O servidor de recursos opera sobre um *Data Warehouse* [1], criado especificamente para atender às demandas dos serviços implementados. Um *Data Warehouse* é, de forma grosseira, um agrupamento de dados provindos de uma ou mais bases possivelmente heterogêneas, que tem por objetivo manter um histórico dos dados além de otimizar as consultas de leitura, aliviando a carga sobre as bases de dados operacionais de um ou mais sistemas. No DataUSP-PósGrad foi criado um modelo multidimensional (cubo OLAP) que agrega diversas bases de outros sistemas da USP, como o *Janus*, o *Juno*, as bases de produção científica (incluindo os dados enviados à *Capes*), entre outros.

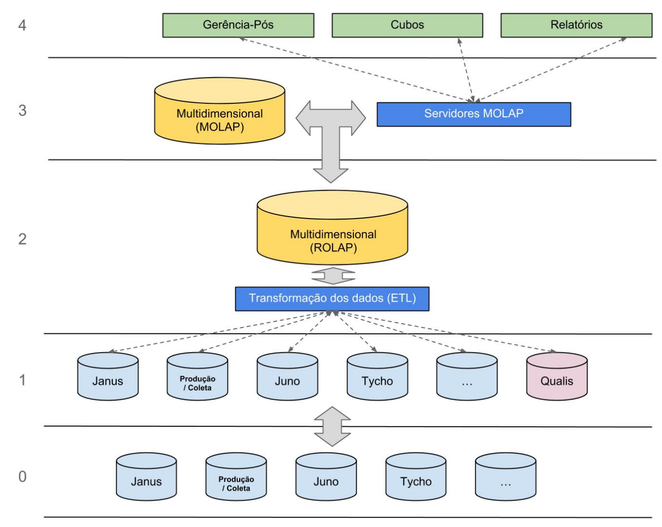

Figura 4 : *Data Warehouse* do DataUSP-PósGrad, com replicação no nível 1 e o multidimensional no nível 3.

Web crawling

Além dos relatórios *Ad hoc*, o DataUSP-PósGrad oferece serviços que disponibilizam informações sobre a quantidade de citações da produção intelectual dos docentes da USP em outros trabalhos. Para isso um outro pequeno sistema semiautônomo, desenvolvido em *Python* [8], varre o site do *Scopus.com* [9] e também o *Google Scholar* [10] para buscar dados de publicações e armazená-los em uma base de dados.

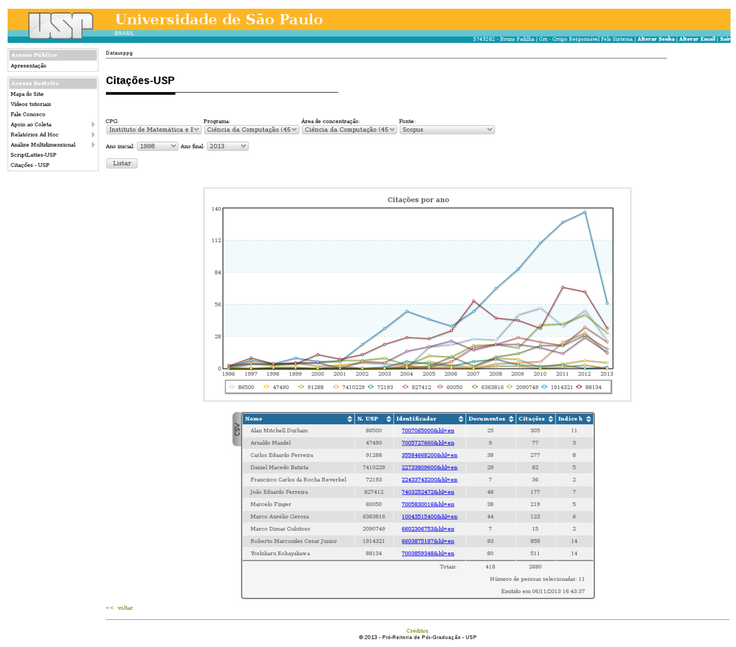

Figura 5 : Gráfico comparativo com o número de citações das publicações de alguns docentes do IME

Conclusão

O sistema teve ótima aceitação entre os dirigentes dos programas de pós-graduação da USP, com matérias publicadas na revista Espaço Aberto da USP n. 151, n. 152 e n. 153, e foi muito elogiado por parte do pró-reitor de pós-graduação Prof. Dr. Vahan Agopyan e, também obteve amplo apoio do Departamento de Informática da USP.
Uma oficina para capacitação dos usuários foi realizada em setembro em todos os Campi da universidade, com uma segunda edição já agendada para o final de Novembro.

Referências

[1] O K Takai, I C Italiano e J E Ferreira,Introdução a Banco de Dados, cp.10 - Data Warehouse, Uma visão geral , *http://www.ime.usp.br/ jef/apostila.pdf*

[2] C Adamson, M Venerable. Data Warehouse Design Solutions, John Wiley & Sons, 1998.

[3] M Golfarelli, D Maio, S Rizzi. "Conceptual Design of Data Warehouses from E/R Schemes", Proceedings of the Hawaii International Conference on System Sciences, Kona, Hawaii, USA, 1998

[4] Roy Fielding, Architectural Styles and the Design of Network-based Software Architectures, 2000, *http://www.ics.uci.edu/ fielding/pubs/dissertation/top.htm*

[5] Jersey, RESTful Web Services in Java, *http://docs.oracle.com/javaee/6/tutorial/doc/giepu.html*

[6] JQuery, *http://jquery.com/*

[7] Ajax, Asynchronous JavaScript and XML, *http://www.w3schools.com/ajax/ajax_intro.asp*

[8] Python, `http://www.python.org/`

[9] Scopus, Citation database, *http://www.elsevier.com/online-tools/scopus*

[10] Google Scholar, *http://scholar.google.com.br/intl/pt-BR/scholar/about.html*