

# Exploração Visual *Online* de Bases de Dados Astronómicas

Tiago G. Mendes<sup>1</sup>, Bruno F.A. Rodrigues<sup>1</sup>, André Moitinho<sup>3</sup>, Miguel Gomes<sup>3</sup>,  
Francisco M. Couto<sup>2</sup>, Tiago J. V. Guerreiro<sup>2</sup>, and Alberto Krone-Martins<sup>3</sup>

<sup>1</sup> Departamento de Engenharia Geográfica, Geofísica e Energia,

<sup>2</sup> LaSIGE, Departamento de Informática,

Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

<sup>3</sup> SIM,

Faculdade de Ciências, Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016  
Lisboa, Portugal

fc33684@alunos.fc.ul.pt, fc37292@alunos.fc.ul.pt, andre@sim.ul.pt,  
mdg@uninova.pt, fcouto@di.fc.ul.pt, tjvg@di.fc.ul.pt, algol@sim.ul.pt

**Resumo** A visualização e exploração de informação nos dias de hoje é de extrema importância para o processo científico. No âmbito da astronomia é essencial existirem ferramentas que possibilitem a exploração de informação e que tenham a capacidade de processamento e visualização de milhões de dados.

Neste artigo é apresentado um protótipo de uma nova aplicação web, que permitirá aos utilizadores de dados astronómicos trabalhar com grandes bases de dados de informação, sobretudo de catálogos de estrelas. Esta aplicação web baseou-se na arquitetura PaaS (*Platform as a Service*), constituída por um *back-end*, *web services* e um *front-end*.

A ferramenta apresentada neste artigo - VEJA (*Visual Exploration Just Anywhere*), é a uma das raras ferramentas web que possibilita a exploração multidimensional de dados de astronomia, através de vários gráficos personalizáveis com vistas ligadas (*linked views*) e com capacidade de trabalhar com múltiplos conjuntos de dados e grande quantidade de informação, sendo a primeira que integra estas potencialidades com bases de dados astronómicas, não havendo necessidade de o utilizador ter de carregar ficheiros com dados para poder visualizar a informação.

**Keywords:** *Big Data*, Astronomia, Visualização, *Linked Views*

## 1 Introdução

A exploração visual de dados astronómicos provenientes de bases de dados disponíveis para o público representa um passo importante para a astronomia. Os enormes arquivos produzidos por observações astronómicas, publicamente disponíveis online pela ESA, NASA, institutos de investigação, consórcios e centros de dados por todo o mundo, em conjunto com os padrões do observatório virtual

criaram um novo paradigma na forma de exploração dos dados. Apesar de já existirem serviços web para interrogar as bases de dados existentes, é ainda necessário para o utilizador descarregar estes dados e explorá-los numa ferramenta local. Existe uma falta de serviços baseados na web que permitam a exploração visual dos dados integrada com as bases de dados atualmente existentes. Tendencialmente os dados deixarão de estar presentes na máquina do utilizador, passando a estar disponíveis a todos através de serviços *online*. A disponibilização e visualização de dados são constituem uma parte do quarto paradigma [5], um tópico fundamental na investigação científica, em particular nos tempos que correm, graças à quantidade enorme de dados disponíveis atualmente.

Os factos acima referidos levaram à necessidade de criar uma ferramenta acessível em qualquer lugar, utilizando um dispositivo qualquer (PC, *tablet* ou *smartphone*) com recurso a *web services* capazes de permitir a exploração desta quantidade de informação.

Uma ferramenta deste tipo deve ter a capacidade de rapidamente permitir a visualização e manipulação de catálogos com milhares de milhões de entradas e dezenas ou centenas de atributos. Atualmente já existem catálogos com cerca de dois milhões de entradas. A missão Gaia<sup>1</sup> da Agência Espacial Europeia (ESA) foi lançada em Dezembro de 2013, esperando-se a recolha de medições relativas a mais de um bilião de estrelas, não apenas na nossa galáxia, mas também noutras, resultando em cerca de um petabyte de informação, e representando o maior conjunto de dados astronómicos até agora.

A Faculdade de Ciências da Universidade de Lisboa está envolvida neste projeto, liderando a participação nacional, sendo um dos objetivos principais a procura, conceção e desenvolvimento deste tipo de serviços de visualização, permitindo a exploração dos seus produtos.

No presente artigo é apresentada uma nova plataforma que permite aos utilizadores de dados astronómicos trabalhar com grandes bases de dados de informação, baseada na arquitetura PaaS (*Platform as a Service*) e constituída por um *back-end*, *web services* e *front-end*.

A aplicação VEJA (*Visual Exploration Just Anywhere*) possibilita a exploração multidimensional de dados através de vários gráficos personalizáveis com vistas ligadas (*linked views*, i.e., a seleção de um ou vários pontos num gráfico é refletida nos restantes) e capacidade de trabalhar com múltiplos conjuntos de dados. A natureza web do serviço VEJA permite que qualquer utilizador, em qualquer parte do mundo com acesso à internet, aceda à informação sem necessidade de instalação de *software*.

Atualmente já existem ferramentas que permitem a exploração de dados na web através de vários gráficos com vistas ligadas (por exemplo a ferramenta *plotly*<sup>2</sup> ou *Filtergraph*<sup>3</sup>). Contudo estas ferramentas não estão direcionadas especificamente para dados astronómicos e requerem que o utilizador tenha os ficheiros no próprio dispositivo e faça o *upload* dos mesmos. Existem também ferramen-

<sup>1</sup> <http://sci.esa.int/gaia/> acedido a 20 de Junho de 2014

<sup>2</sup> <https://plot.ly/> acedido a 16 de Julho de 2014

<sup>3</sup> <http://filtergraph.vanderbilt.edu/> acedido a 16 de Julho de 2014

tas desenvolvidas especificamente para a comunidade astronómica (por exemplo, TOPCAT<sup>4</sup>) não sendo, contudo, baseadas na web. Esta é a primeira ferramenta que surge como resposta ao desafio de recolha, armazenamento, disponibilização e exploração visual *online* eficaz de um grande volume de dados obtidos em missões de astronomia, ligados entre si, esperando-se que com futuros desenvolvimentos seja possível a análise visual de conjuntos de dados com biliões de objetos. Deste modo a ferramenta VEJA pode vir a implantar-se como uma referência na exploração de dados de astronomia, servindo as futuras missões de observação (por exemplo a Gaia). Atualmente existe informação relativa a 27348 catálogos provenientes de uma fonte de dados, sendo que alguns contêm cerca de dois milhões de entradas.

No presente artigo é feita, em primeiro lugar, uma descrição da arquitetura implementada, seguindo-se da descrição do *back-end*, onde é explicado o acesso, tratamento e estrutura de dados, assim como as tecnologias utilizadas. Segue-se a descrição dos *web services* e a forma de disponibilização dos dados e a descrição do *front-end*, que constitui a interface com utilizador. Por fim é discutido o trabalho futuro a ser desenvolvido no âmbito da ferramenta VEJA.

## 2 Arquitetura

A arquitetura implementada encontra-se esquematizada na Figura 1. O acesso aos dados é feito por HTTP ao serviço VizieR [1] (descrito adiante) e, de seguida, guardado na base de dados MySQL ou em disco, constituindo o *back-end*. Estes dados são disponibilizados em REST (*Representational State Transfer*), permitindo o acesso HTTP. Por fim, os dados são utilizados pelo *front-end* para exploração e disponibilização gráfica da informação.

## 3 *Back-end*

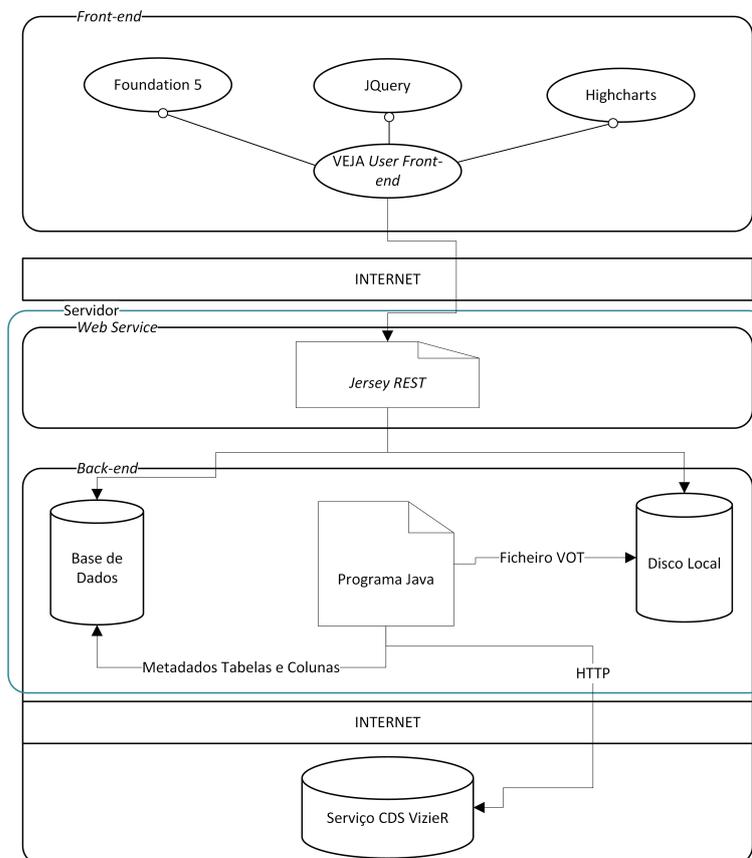
### 3.1 Dados e Metadados

Os dados diretamente provenientes das fontes de informação, através de interrogação direta aos serviços web (ver secção 3.3), são de três tipos:

- Metadados das tabelas (que correspondem a catálogos provenientes de observações astronómicas)
- Metadados das colunas
- Dados propriamente ditos

Os Metadados das tabelas contêm uma descrição sumária dos dados contidos no catálogo. Cada tabela ou catálogo é constituída por um número variável de colunas, dependente do catálogo escolhido. Os Metadados relativos a essas colunas incluem o nome da coluna, uma breve descrição da coluna, o tipo de dados da coluna (e.g. *double*, *char*, etc.), e a unidade da coluna (e.g. mm). Estes Metadados

<sup>4</sup> <http://www.star.bris.ac.uk/~mbt/topcat/> acedido a 16 de Julho de 2014



**Figura 1.** Arquitetura implementada na plataforma VEJA

permitirão ao utilizador filtrar a informação sobre a qual pretende trabalhar, não sendo necessário descarregar os dados todos de um catálogo. Atualmente existe informação relativa a 27348 catálogos guardada na base de dados. Por fim, os dados propriamente ditos dizem respeito aos exemplares (linhas) de cada tabela. Atualmente existem tabelas que podem conter cerca de dois milhões de linhas, representando um grande volume de informação. Para além da informação de base acima descrita, optou-se por recolher informação adicional contendo descrições mais pormenorizadas dos catálogos e nomes dos autores dos catálogos, melhorando a experiência para o utilizador. Os metadados (relativos a catálogos e colunas) são gravados numa base de dados MySQL, sendo os dados em si guardados num ficheiro em disco (como descrito adiante).

### 3.2 Tecnologias Utilizadas

**Linguagens de Programação** O *back-end* foi implementado com vista a uma melhor integração com ferramentas de acesso e manipulação de dados astronómicos já existentes. Estas incluem as APIs (*Application Programming Interface*) STIL (*Starlink Tables Infrastructure Library*) [6] e STILTS (*Starlink Tables Infrastructure Library Tool Set*) [7], que facilitam o acesso às fontes de dados e a manipulação e análise sintática dos dados (explicada de seguida). Dado que estas ferramentas foram desenvolvidas em JAVA, optou-se por usar esta linguagem de programação na aplicação VEJA.

**Bibliotecas Auxiliares** Como referido acima, foram utilizadas duas bibliotecas auxiliares, STIL e STILTS. Estas bibliotecas facilitam o acesso aos dados presentes nas bases de dados astronómicas, cujo acesso é efetuado através um protocolo recomendado pelo IVOA (*International Virtual Observatory Alliance*) designado por TAP (*Table Access Protocol*) [2]. Para além disso, estas ferramentas permitem a análise sintática dos dados em formato VOTable [3], um formato semelhante ao formato XML utilizado pela comunidade astronómica, para que os mesmos possam ser guardados como valores na base de dados. As ferramentas possibilitam também efetuar diversas operações sobre os dados, como por exemplo operações matemáticas (adição entre colunas, entre uma coluna e um número, etc.), embora tal não tenha sido explorado na presente aplicação.

### 3.3 Fontes de Informação

Para a fase inicial da aplicação VEJA optou-se por efetuar o acesso apenas a uma fonte de informação - CDS (*Centre de Données astronomiques de Strasbourg*) *VizieR Astronomical Database* [1]. Como mencionado acima, o acesso é feito por um protocolo denominado TAP. Este protocolo surge de uma recomendação do IVOA e permite o acesso a dados tabulares (incluindo dados astronómicos), à base de dados CDS, aos Metadados das tabelas e aos dados propriamente ditos. Para além disso, permite o acesso através de diferentes linguagens de interrogação, que inclui a linguagem ADQL - *Astronomical Data Query Language* [4]. O protocolo TAP permite o acesso aos dados como um serviço web, através de operações HTTP.

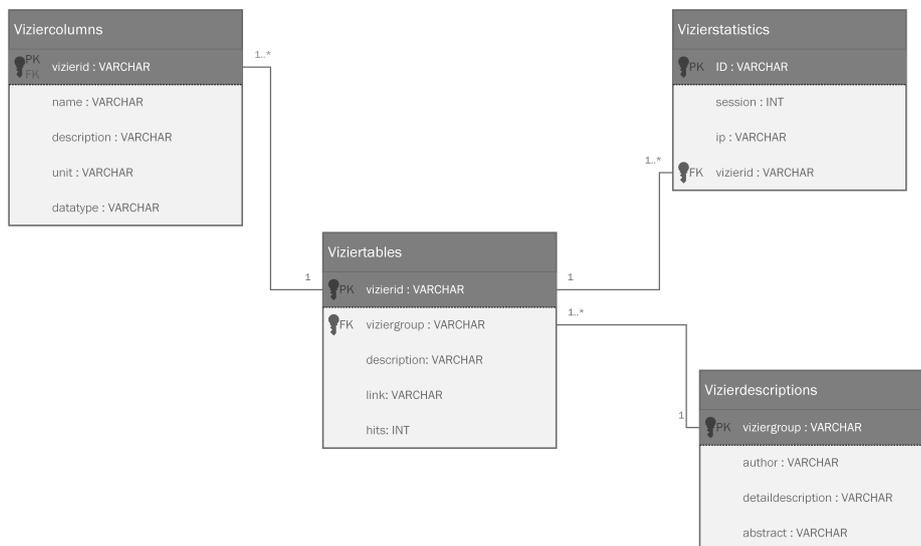
Dado que o serviço VizieR implementa o protocolo TAP, é possível aceder tanto aos dados como aos Metadados de tabelas e colunas. A resposta é devolvida no formato VOTable. Os dados obtidos após a interrogação foram analisados sintaticamente através do STIL e posteriormente gravados numa base de dados própria.

### 3.4 Estrutura de Dados

Visto que a quantidade de dados existentes nas bases de dados astronómicas é imensa, optou-se por adotar uma estratégia de otimização de informação (Figura 2), que consistiu em guardar na base de dados de raiz apenas os Metadados

das tabelas, o que permite aos utilizadores verem as tabelas existentes e as suas descrições. Caso um utilizador selecione uma tabela sobre a qual pretenda trabalhar é feito o acesso às fontes de informação e descarregado o ficheiro de Metadados das colunas, que fica gravado na base de dados. Simultaneamente são descarregados os dados do catálogo, que ficam gravados num ficheiro de texto, comprimido, em disco. Deste modo a base de dados é preenchida *on-demand*, evitando sobrecarregá-la com ficheiros que nunca são utilizados.

A opção de gravar os dados em ficheiro de texto prende-se com a rapidez de análise sintática das ferramentas STIL e STILTS, que se revela mais eficiente que a interrogação de dados diretamente à base de dados.



**Figura 2.** Esquema UML da base de dados implementada

A tabela “Viziertables” contém os Metadados das tabelas, identificadas pela chave primária “vzierid”. Cada tabela é um subgrupo de um catálogo, identificado pela chave estrangeira “vziergroup”, utilizada para referenciar a tabela de descrições mais detalhadas “Vizierdescriptions”. Para além das colunas acima referidas, existe também uma coluna para a descrição dos dados, uma coluna com um link para a localização em disco do ficheiro VOTable com os dados e ainda uma coluna que grava o número de vezes que a tabela foi selecionada em *front-end*.

A tabela “Vizierdescriptions” contém o autor do catálogo e as descrições e resumos detalhados.

A tabela “Viziercolumns” contém os Metadados das colunas de uma tabela, descritos na Secção 3.1.

Por fim, a tabela “Vizierstatistics” guarda estatísticas associadas à utilização das tabelas por parte dos utilizadores. Quando um utilizador seleciona uma tabela é registado o IP desse utilizador, a tabela selecionada e a data de seleção. Uma sessão corresponde a seleções feitas por um mesmo IP numa mesma data. Isto permite que sejam guardadas estatísticas relativamente à utilização de tabelas por parte dos utilizadores. Deste modo são recolhidas estatísticas de tabelas utilizadas pelo mesmo utilizador que podem ser utilizadas num sistema de sugestão de tabelas (descrito na Secção 5).

## 4 *Web Service*

### 4.1 *Tecnologias Utilizadas*

O *web service* foi desenvolvido em JAVA através da *framework* Jersey, uma implementação da API JAX-RS. Esta *framework* facilitou a implementação do *web service* REST, através de vários métodos para acesso aos dados gravados em *back-end*. Após serem feitos os pedidos HTTP, a resposta é devolvida em formato JSON (*Javascript Object Notation*), que facilita o uso dos dados em *front-end* através da linguagem Javascript. A conversão de objetos JAVA para JSON é feita através da API Jackson, que converte automaticamente os objetos.

### 4.2 *Lista de Métodos*

Na Tabela 4.2 é apresentada uma lista de métodos e os *endpoints* (endereços) implementados que permitem o acesso aos dados gravados em *back-end*.

## 5 *Front-end*

### 5.1 *VEJA - Visual Exploration Just Anywhere*

O VEJA é uma aplicação web 2.0 que pretende focar-se no utilizador e servir de plataforma dinâmica e colaborativa, tendo sido implementadas algumas funcionalidades, para além da visualização gráfica, que incluem a capacidade de recolher *feedback* dos catálogos mais explorados e a recomendação de tabelas com base nas escolhas de outros utilizadores. Outras funcionalidades serão implementadas no futuro, e encontram-se descritas na secção 7.

O VEJA foi implementado para ser de fácil utilização e permitir a visualização de múltiplos gráficos em simultâneo, provenientes de diferentes catálogos, e com capacidade para oferecer vistas ligadas (*linked views*). As vistas ligadas permitem que a seleção de um ou vários pontos num gráfico seja refletida em todos os restantes gráficos, o que significa que, para um mesmo catálogo, seja possível criar múltiplos gráficos com atributos diferentes nos eixos onde os dados estão sempre ligados. A necessidade das vistas ligadas surge pelo facto de, para um

**Tabela 1.** Dados disponibilizados pelo *web service* e respetivos *endpoints*

URI	Método HTTP	Parâmetros	Resultado
/table/top	GET	-	Obter tabelas mais populares.
/table/filter/startpoint/increment	GET	Tabela, ponto inicial e número de linhas	Obter tabela pesquisada.
/columns/catalog	GET	Tabela pretendida	Obter metadados de colunas de uma tabela.
/votdata	POST	Tabela pretendida, lista de colunas pretendidas e número máximo de linhas	Obter dados de catálogo.
/statistics/hits	POST	ID da tabela	Adicionar tabela à contagem de tabelas mais populares.
/table/suggest/vid	GET	ID da tabela	Obter outras tabelas relacionadas
/statistics/user	POST	IP e ID da tabela	Adicionar estatística de utilizador

mesmo catálogo, existirem diversas colunas diferentes sendo possível para o utilizador criar dois ou mais gráficos do mesmo catálogo, com vistas de diferentes colunas, e ver o ponto correspondente a um dado objeto selecionado em todos os gráficos. Esta capacidade tem um potencial e utilidade enormes na exploração de dados astronómicos.

## 5.2 Tecnologias Utilizadas

O *front-end* foi desenvolvido tendo como base a *framework* Foundation 5<sup>5</sup>, que permite o desenvolvimento eficaz de *websites* com classes CSS, funcionando em modo de grelha (onde a disposição das componentes HTML é feita como uma tabela, ao longo de várias colunas e linhas). Este pormenor permite criar aplicações que se adaptam a qualquer dispositivo.

Foi também utilizada a biblioteca jQuery<sup>6</sup>, que facilitou o acesso ao *web service* através de pedidos AJAX (*Asynchronous Javascript and XML*), manipulação de DOM (*Document Object Model*) e animação de páginas HTML.

A visualização de gráficos foi implementada através da biblioteca Highcharts<sup>7</sup>, desenvolvida em JavaScript, que permite a criação, manipulação e personalização de diversos tipos de gráficos de forma eficaz.

## 5.3 Funcionamento do VEJA

O VEJA permite ao utilizador pesquisar por tabelas presentes na base de dados (Figura 3) ou escolher uma tabela entre as tabelas mais vistas. Como referido anteriormente, apenas os Metadados de tabelas estão presentes originalmente, sendo necessário um utilizador selecionar uma tabela para descarregar a restante informação. Após selecionar uma tabela segue-se a seleção das colunas sobre as quais o utilizador pretenda trabalhar, assim como o número de linhas a descarregar (caso o utilizador pretenda explorar apenas uma parte dos dados) (Figura 4). Para além disso são enviados dados para efeitos estatísticos para o *back-end* com informação da tabela selecionada, data em que foi selecionada e IP do utilizador.

Após a seleção das colunas, é feita (em *back-end*) uma análise sintática do ficheiro VOTable gravado em disco, sendo posteriormente convertido num objeto em JSON. Do ficheiro total é apenas enviada como resposta a parte relevante para o utilizador (colunas escolhidas e número de linhas), ou seja, a informação devolvida é personalizada às necessidades do utilizador. O acesso aos dados é feito através do *web service* descrito na Secção 4.

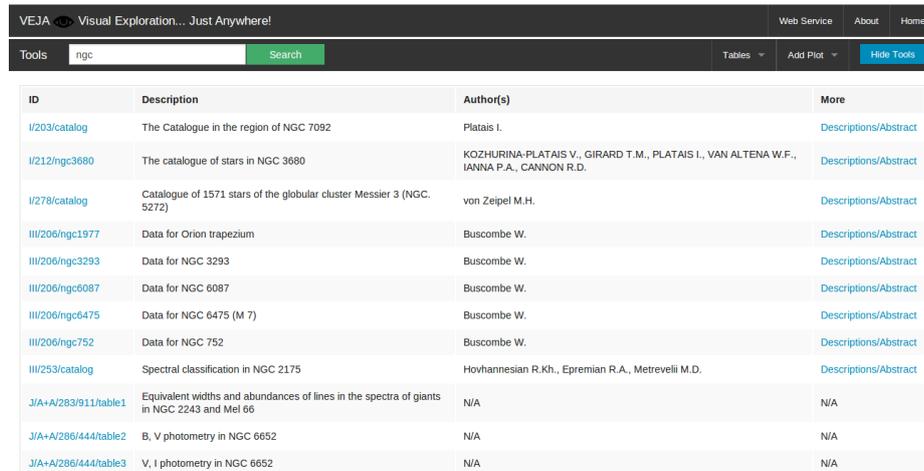
Os dados da resposta são gravados num objeto JavaScript, podendo ser utilizados posteriormente. Após estas operações, o utilizador tem a opção de ver tabelas relacionadas (tabelas que outros utilizadores tenham selecionado quando também selecionaram a tabela a ser explorada), personalizar um gráfico ou adicionar mais gráficos em linha ou coluna. Esta funcionalidade permite personalizar a área de trabalho, podendo ter até quatro gráficos em coluna e dez gráficos em linha.

Os gráficos podem ser personalizados de diversas formas (Figura 5). Os pontos dos gráficos podem ter cores e formas diferentes, é possível escolher que colunas associar ao eixo XX e YY e ainda atribuir um título ao gráfico e aos

<sup>5</sup> <http://foundation.zurb.com/> acedido a 20 de Junho de 2014

<sup>6</sup> <http://jquery.com/> acedido a 20 de Junho de 2014

<sup>7</sup> <http://www.highcharts.com/> acedido a 20 de Junho de 2014

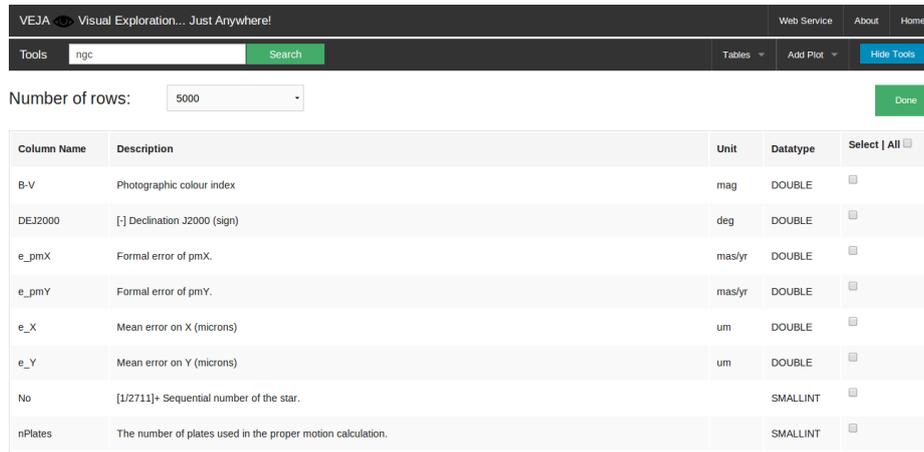


VEJA Visual Exploration... Just Anywhere! Web Service About Home

Tools ngc Search Tables Add Plot Hide Tools

ID	Description	Author(s)	More
I/203/catalog	The Catalogue in the region of NGC 7092	Platais I.	<a href="#">Descriptions/Abstract</a>
I/212/ngc3680	The catalogue of stars in NGC 3680	KOZHURINA-PLATAIS V., GIRARD T.M., PLATAIS I., VAN ALTEN A W.F., IANNA P.A., CANNON R.D.	<a href="#">Descriptions/Abstract</a>
I/278/catalog	Catalogue of 1571 stars of the globular cluster Messier 3 (NGC 5272)	von Zeipel M.H.	<a href="#">Descriptions/Abstract</a>
III/206/ngc1977	Data for Orion trapezium	Buscombe W.	<a href="#">Descriptions/Abstract</a>
III/206/ngc3293	Data for NGC 3293	Buscombe W.	<a href="#">Descriptions/Abstract</a>
III/206/ngc6087	Data for NGC 6087	Buscombe W.	<a href="#">Descriptions/Abstract</a>
III/206/ngc6475	Data for NGC 6475 (M 7)	Buscombe W.	<a href="#">Descriptions/Abstract</a>
III/206/ngc752	Data for NGC 752	Buscombe W.	<a href="#">Descriptions/Abstract</a>
III/253/catalog	Spectral classification in NGC 2175	Hovhannesian R.Kh., Epremian R.A., Metrevelli M.D.	<a href="#">Descriptions/Abstract</a>
J/A+A/283/911/table1	Equivalent widths and abundances of lines in the spectra of giants in NGC 2243 and Mel 66	N/A	N/A
J/A+A/286/444/table2	B, V photometry in NGC 6652	N/A	N/A
J/A+A/286/444/table3	V, I photometry in NGC 6652	N/A	N/A

**Figura 3.** Seleção de tabelas para exploração



VEJA Visual Exploration... Just Anywhere! Web Service About Home

Tools ngc Search Tables Add Plot Hide Tools

Number of rows: 5000 Done

Column Name	Description	Unit	Datatype	Select   All
B-V	Photographic colour index	mag	DOUBLE	<input type="checkbox"/>
DEJ2000	[ ] Declination J2000 (sign)	deg	DOUBLE	<input type="checkbox"/>
e_pmX	Formal error of pmX.	mas/yr	DOUBLE	<input type="checkbox"/>
e_pmY	Formal error of pmY.	mas/yr	DOUBLE	<input type="checkbox"/>
e_X	Mean error on X (microns)	um	DOUBLE	<input type="checkbox"/>
e_Y	Mean error on Y (microns)	um	DOUBLE	<input type="checkbox"/>
No	[1/2711]+ Sequential number of the star.		SMALLINT	<input type="checkbox"/>
nPlates	The number of plates used in the proper motion calculation.		SMALLINT	<input type="checkbox"/>

**Figura 4.** Seleção de colunas

eixos. É ainda possível inverter os eixos, um passo importante na exploração de dados astronômicos. O resultado final da visualização é apresentado na Figura 6.

## 6 Trabalho Futuro

Apesar de algumas limitações da aplicação VEJA no presente (relativamente á quantidade de dados com que é possível trabalhar e às ainda limitadas funcionalidades como ferramenta colaborativa), a natureza dinâmica da web permite uma atualização constante e simples das aplicações em geral, sendo por isso fácil

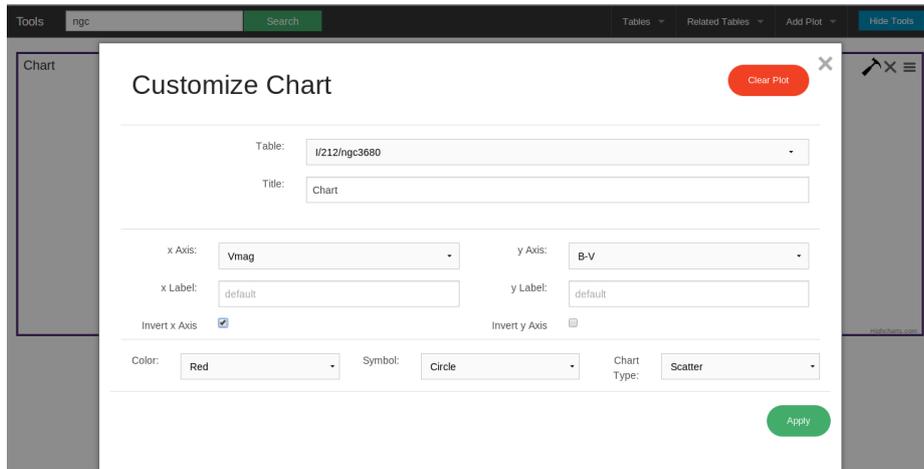


Figura 5. Personalização de gráficos

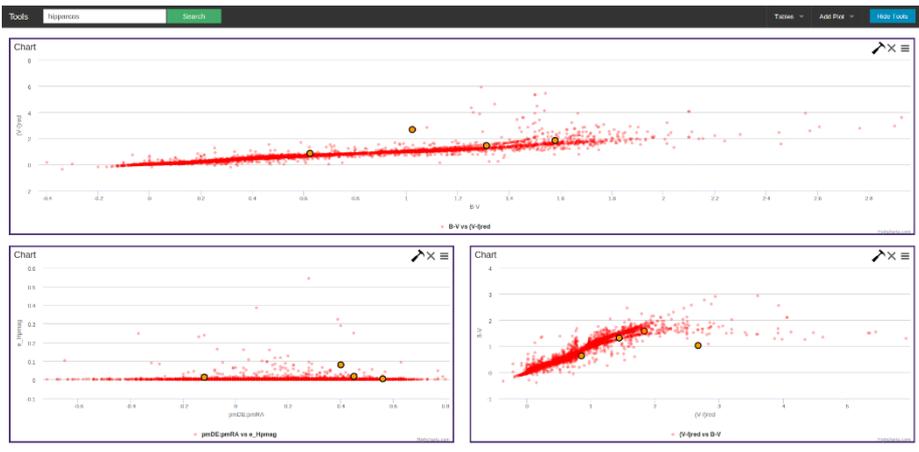


Figura 6. Múltiplos gráficos visualizados em simultâneo com vistas ligadas. O ponto selecionado num gráfico é automaticamente selecionado nos restantes gráficos.

introduzir novas ferramentas e funcionalidades instantaneamente sem necessidade de atualizações em pacotes de *software*. A aplicação, tal como apresentada, fez parte de uma fase piloto para comprovar com sucesso o conceito. Esta fase abriu portas para o desenvolvimento de uma aplicação mais robusta e capaz, que permitirá servir toda a comunidade astronómica.

O próximo passo a implementar consiste tornar a aplicação mais colaborativa, que se revela bastante importante para a comunidade astronómica. Neste âmbito, está pensada a implementação de um sistema de registo e partilha de operações efetuadas sobre os gráficos. Isto permitirá a um utilizador não só retomar o seu

trabalho como também partilhar o seu trabalho com outros utilizadores, algo fundamental no meio astronómico. Outra ideia considerada é a criação de uma lista de catálogos favoritos, também personalizados pelo utilizador. Estes passos serão feitos após introdução de um sistema de contas de utilizadores.

Pretende-se também que a aplicação venha a abranger mais fontes de dados de centros que forneçam serviços do IVOA, não sendo limitado ao CDS, e servindo como centralização da informação existente na web.

Um dos maiores problemas associados à quantidade de dados existentes consiste na sobrecarga induzida aos *browsers* na exploração da informação. É fácil perceber que a visualização gráfica de um milhão de pontos num browser é uma sobrecarga elevadíssima. Isto requer que seja implementada uma forma de otimização dos dados. Em primeiro lugar é necessário ter em conta que, para o utilizador, a visualização de um milhão de pontos num browser constitui um excesso de informação enorme e sem significado, pois não será possível distinguir esses pontos todos. Assim, será necessário criar níveis de abstração e generalização dos dados, com base na posição destes. Este processo já se encontra em desenvolvimento, podendo ser implementado através de árvores. Deste modo, a visualização original consiste apenas num pequeno subconjunto dos dados. À medida que o utilizador amplia uma zona de interesse vai havendo uma densificação dos dados. Isto é possível graças aos pedidos AJAX, permitindo que tudo se processe em tempo quase real sem sobrecarregar o browser. Este modo de funcionamento introduz uma alteração fundamental na aplicação - em vez de carregar de uma só vez todos os dados que o utilizador pretende, esses dados vão sendo carregados à medida das necessidades. Neste âmbito, a aplicação revelou problemas de performance ao trabalhar com 100,000 ou mais linhas de dados.

Para além da otimização da visualização dos dados, haverá também suporte para grandes bases de dados, baseado na integração com o *software* servidor de visualizações que se encontra em desenvolvimento no âmbito da missão Gaia.

Outras modificações incluem pequenas alterações no modo de visualização dos dados, a possibilidade de efetuar operações sobre os dados (como somas, subtrações, divisões e multiplicações entre colunas e entre colunas e números), a capacidade de importação de ficheiros locais (no disco do utilizador), entre outras.

## 7 Conclusões

No presente artigo foi apresentado um protótipo de uma aplicação que surge como resposta face ao desafio da visualização e gestão de informação proveniente de missões de observação astronómica, que tendencialmente recolhem volumes cada vez maiores de informação. A ferramenta VEJA é um primeiro passo para responder às necessidades da comunidade astronómica no que diz respeito à análise visual da informação. Foi demonstrada a capacidade da web, atualmente, fazer face a estes desafios, sobretudo com o surgimento do HTML5 e AJAX, que permitem introduzir um elevado dinamismo às páginas web sem necessidade de instalação adicional de ferramentas. Conjugado com a capacidade de, atu-

almente, maior parte dos *browsers* funcionar com as diversas APIs existentes possibilitou o desenvolvimento de uma aplicação capaz de chegar a todos os utilizadores. Apesar da existência de outras ferramentas que permitem a exploração de dados na web, a VEJA é a primeira aplicação que integra a visualização com bases de dados de astronomia.

A web permite, atualmente, a criação de ferramentas potentes disponíveis para utilizadores a nível global. As limitações da web para este tipo de aplicações residem na memória e processamento no âmbito da visualização e manipulação de grandes quantidades de informação. Contudo, técnicas de otimização aplicadas aos dados permitem ultrapassar estes problemas.

Espera-se que com mais desenvolvimentos (no que diz respeito à escalabilidade, eficácia na gestão de grandes conjuntos de dados e melhorias no sentido de tornar a plataforma mais colaborativa) a aplicação VEJA se implante na comunidade astronómica como uma ferramenta de referência na exploração visual na web de conjuntos de dados provenientes de observações astronómicas e de vários centros de dados.

## 8 Agradecimentos

Este projeto foi desenvolvido com o apoio da FCT Portuguesa para o Projeto Estratégico LaSIGE, ref. PEst-OE/EEI/UI0408/2014 e do Laboratório de Sistemas, Instrumentação e Modelação em Ciências e Tecnologias do Ambiente e do Espaço (SIM).

## Referências

1. CDS, Strasbourg, France (2014) VizieR Catalogue Access Tool . <http://vizier.u-strasbg.fr/viz-bin/VizieR>. Acedido a 25 Maio 2014
2. Dowler, P., Rixon, D., Tody, D. (2010) Table Access Protocol. International Virtual Observatory Alliance. Disponível em <http://www.ivoa.net/documents/TAP/>. Acedido a 25 Maio 2014
3. Ochsenbein, F., Williams, R., Davenhall, C., Demleitner, M., Durand, D., Fernique, P., Giarretta, D., Hanisch, R., McGlynn, T., Szalay, A., Taylor, M., Wicenec, A. (2013) VOTable Format Definition. International Virtual Observatory Alliance. Disponível em <http://www.ivoa.net/documents/VOTable/>. Acedido a 25 Maio 2014
4. Ortiz, I., Lusted, J., Dowler, P., Szalay, A., Shirasaki, Y., Nieto-Santisteban, M.A., Ohishi, M., OMullane, W., Osuna, P., VOQL-TEG, VOQL Working Group (2008) Astronomical Data Query Language. International Virtual Observatory Alliance. Disponível em <http://www.ivoa.net/documents/latest/ADQL.html>. Acedido a 25 Maio 2014
5. Hey, T., Tansley, S., Tolle, K. (2009) The Fourth Paradigm, Data-Intensive Scientific Discovery. Microsoft, United States of America.
6. Taylor, M. (2013) Starlink User Note252. Starlink. Disponível em <http://www.star.bristol.ac.uk/mbt/stil/sun252.pdf>. Acedido a 25 Maio 2014
7. Taylor, M. (2013) Starlink User Note256. Starlink. Disponível em <http://www.star.bris.ac.uk/mbt/stilts/sun256.pdf>. Acedido a 25 Maio 2014