

P.PORTO <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Enunciado Trabalho Prático: Parte 2	Ano letivo 2018/2019	Data
	Curso Licenciatura em Engenharia Informática	Hora	
	Unidade Curricular Processamento Estruturado de Informação	Duração	

Destinatários

Este trabalho destina-se a todos os estudantes inscritos na unidade curricular de Processamento Estruturado de Informação que pretendam obter aprovação à Unidade Curricular.

Objetivos

Com a realização do trabalho prático, pretende-se que os alunos ponham em prática todos os conhecimentos adquiridos, demonstrando as suas competências em:

- Conhecer conceitos de armazenamento de documentos e as tecnologias atuais para a estruturação de informação;
- Reconhecer e compreender a semântica e a sintaxe da notação JSON (Javascript Object Notation) bem como mecanismos para a sua preservação e exploração utilizando coleções de documentos;
- Conhecer e implementar web APIs para suportar o processo de preservação, processamento e exploração de documentos.

Enunciado

Com o objetivo de obter uma visão integrada dos dados de vendas de produtos associados a cada loja, a BikeOnTrack iniciou recentemente um projeto que visa suportar de forma mais eficaz os processos de tomada de decisão relacionados com os produtos comercializados. Estes novos processos pressupõem que cada loja comunique mensalmente toda a informação relacionada com as vendas realizadas. Numa primeira fase do projeto foi desenvolvido um vocabulário XML para suportar o armazenamento de dados em documentos de auditoria. A BikeOnTrack pretende agora iniciar os esforços necessários para que os dados extraídos diretamente dos sistemas de faturação de cada loja possam ser alvo de um processo de processamento e preservação utilizando um repositório específico para que possa ser utilizada a geração automática de documentos XML suportados pelo vocabulário desenvolvido.

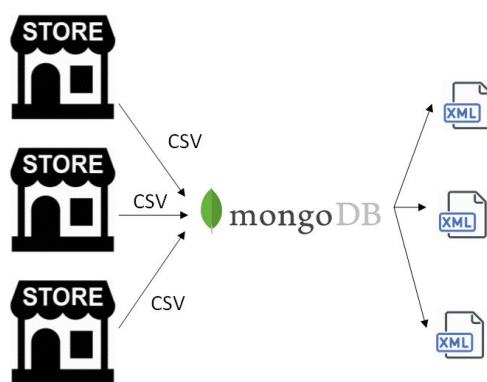


Figura 1

Foram por isso identificados os seguintes pressupostos para a segunda etapa do projeto (Figura 1):

- De forma a não perturbar o normal funcionamento dos sistemas utilizados em cada loja, os dados de cada um dos sistemas de faturação devem ser extraídos (sem qualquer tipo de pré-tratamento) para o formato CSV para posterior importação numa base de dados em orientada a documentos (por exemplo, MongoDB);

P.PORTO <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Enunciado Trabalho Prático: Parte 2	Ano letivo 2018/2019	Data
	Curso Licenciatura em Engenharia Informática	Hora	
	Unidade Curricular Processamento Estruturado de Informação	Duração	

- Assim que todas as lojas enviam os seus documentos CSV, os dados armazenados no MongoDB estão preparados para a sua extração para XML de acordo com as regras definidas pelo vocabulário;
- Deverá ser desenvolvido um serviço REST capaz de expor a funcionalidade de exportação dos dados para XML. A invocação do serviço pressupõe a indicação da loja da qual se pretende gerar o documento XML assim como o período a que corresponde o exercício (mês).

São disponibilizados os seguintes documentos de exemplo:


- ProductDetails: Detalhes dos produtos comercializados nas lojas
- ProductListPriceHistory: Histórico dos preços dos produtos
- CurrencyDetails: valor das taxas de câmbio utilizadas nas vendas (SalesDetails)
- SalesDetails: Dados das vendas

Adicionalmente, foram definidas as seguintes regras/alterações para suportar exportação dos dados para XML:

- Os dados dos **produtos (ProductDetails)** e dos **valores de câmbio (CurrencyDetails)** devem estar relacionados com os dados das vendas. Isto é, o documento XML extraído apenas deverá possuir informação dos produtos e valores de câmbio dos produtos/valores de câmbio presentes nas vendas realizadas para um dado período.
- O vocabulário deverá ser enriquecido com os seguintes dados relacionados **com cada venda**:
 - Número total de produtos
 - Número total de produtos diferentes
 - Valor médio do preço de venda dos produtos
- O vocabulário deverá ser enriquecido com os seguintes dados relacionados **com o exercício (conjunto de vendas)**:
 - Número total de produtos
 - Número total de produtos diferentes
 - Número total de clientes diferentes
 - Valor vendido por cada cliente. Ordenar o valor vendido por ordem decrescente
 - Por produto, apresentar o total de unidade vendidas. Os resultados devem estar ordenados de forma descendente pelo número de unidades vendidas
 - Valor total da venda por cada moeda utilizada (considere a moeda: USD para os valores: NULL)
- Deve ainda ser possível incluir um novo conjunto de informação no documento XML gerado que apresente:
 - Total de produtos vendidos por loja
 - Valor total das vendas por loja
 - Valor médio do preço de venda dos produtos por loja
- A manipulação dos dados deverá ser realizada através de consultas MongoDB (utilizando a função `find()` ou `aggregate()`).

É ainda necessário proceder a algumas correções aplicadas aos dados extraídos das fontes de dados:

- Os produtos são comuns a todas as lojas. No entanto, os dados dos produtos (**ProductDetails**) contêm uma incoerência relacionada com o valor de venda (ListPrice) que se encontra com o valor 0. Para isso é disponibilizado um novo conjunto de dados que contêm o histórico do preço dos produtos (**ProductListPriceHistory**). Como tal, este conjunto de dados deve ser utilizado para corrigir o valor do campo ListPrice dos dados dos produtos para o valor que sinalizado nos dados: **ProductListPriceHistory**. No caso de não ser possível definir o valor correto do preço

 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Enunciado Trabalho Prático: Parte 2	Ano letivo 2018/2019	Data
	Curso Licenciatura em Engenharia Informática	Hora	
	Unidade Curricular Processamento Estruturado de Informação	Duração	

(diferente de 0), o produto deverá ser enriquecido com um campo sinalizando a necessidade de correção posterior.

Cada grupo deverá no mínimo desenvolver:

- As alterações necessárias para suportar os requisitos que devem ser adicionados ao vocabulário;
- Um documento com as alterações realizadas no vocabulário para responder aos novos requisitos ou possíveis correções realizadas.
- Um conjunto de XML schemas capazes de validar todas as regras de sintaxe definidas para a linguagem e tipos associados;
- Documentos XML gerados a partir dos dados armazenados na base de dados MongoDB. Um programa em linguagem Java deve ser desenvolvido para suportar a geração dos documentos XML (que representam os dados de auditoria de cada loja) de forma automatizada. A interação com o programa deverá ser realizada utilizando um serviço REST. Deverá ser possível extrair os documentos tendo em consideração uma loja e um período.
- Um **relatório** pormenorizado onde deverá identificar:
 - Os requisitos que levaram às alterações do vocabulário. Para os requisitos que considerar ambíguos, deve justificar a abordagem seguida.
 - Identificar a abordagem seguida para o armazenamento dos dados na base de dados MongoDB, assim como todos os procedimentos/consultas utilizados para manipular/adaptar os dados inseridos.
 - Descrição e documentação dos serviços implementados.
 - Apreciação crítica do trabalho desenvolvido. O relatório representa uma componente **importante (30%)** desta componente.

Nota: O trabalho realizado nesta segunda entrega programada representa uma parte do projeto a desenvolver em subseqüentes entregas. Como tal, o projeto submetido nesta fase poderá ser alterado/melhorado na entrega final do projeto.

Este trabalho deverá ser realizado em grupo composto por **três** alunos (já constituído na primeira entrega do trabalho)

Datas e considerações

O trabalho deve ser entregue até às **23:55** (hora moodle) horas do dia **10 de janeiro de 2019**, devendo a entrega ser feita através da página da unidade curricular de Processamento Estruturado de Informação em <http://moodle.estgf.ipp.pt>.

Observações:

A entrega dos trabalhos deve obedecer aos requisitos identificados no ponto **Formato da entrega**. Os alunos que não realizem a entrega do trabalho até à data/hora definida serão sujeitos a penalização.

Formato da entrega

Os trabalhos entregues deverão evitar (se possível) utilizar caminhos absolutos ou endereços específicos, de modo a que possam ser facilmente utilizados em qualquer máquina. Para além disso, e no sentido de facilitar a receção dos vários trabalhos recebidos, estes deverão observar as seguintes regras:

- **Todos os elementos do grupo** deverão submeter o trabalho no link respetivo;
- O trabalho desenvolvido deverá ser entregue através do moodle, através da submissão de um ficheiro com o nome PEI_EE_<nr_do_aluno>_<nr_do_aluno>.zip.