

## Objetivo geral

O objetivo do estudo de caso é desenvolver um sistema web em 3 camadas que colete automaticamente dados de sites de comércio eletrônico e faça a integração dos dados, indicando propriedades de produtos, fabricantes, evolução de preços e de avaliações, entre outros.

## Descrição

O sistema deve ter um cadastro de websites de importação (e.g., seções de lojas virtuais), que deve incluir a identificação e dados básicos do website, e o conjunto de scripts de crawling para o website. Deve ser mantida a data e hora de inclusão ou alteração de cada script, mantendo-se no banco de dados o histórico de versões de cada script.

Os scripts devem coletar automaticamente dados de categorias de produtos selecionadas. Defina um subconjunto de produtos oferecidos pelas lojas (e.g., uma ou mais categorias de produtos). Os scripts deverão buscar dados dos produtos (descrição, preço etc.) e de entidades associadas (propriedades/especificações técnicas, dados de fabricantes/fornecedores, avaliações, dados de avaliadores como cidade, sexo, faixa etária etc.). Espera-se que os dados coletados possibilitem produzir um modelo de dados com algumas entidades associadas. É necessário que sejam considerados, pelo menos, três sites concorrentes. Deve haver interseção entre os produtos fornecidos pelos diferentes sites. Isto é, pode haver produtos oferecidos por apenas um dos sites, mas é fundamental que existam vários produtos que são oferecidos por dois ou mais sites.

A interface de cadastro de websites deve listar todos os dados de cada website, incluindo histórico de scripts e execuções. A partir desta interface, o usuário faz a solicitação de uma "execução" de crawling. O crawling pode ser invocado via sistema ou realizado anteriormente offline. Os controllers de execução devem processar os arquivos resultantes do crawling, passando-se como parâmetro o nome do arquivo ou invocando o processo de crawling e processando a saída gerada. Os controllers devem invocar os métodos dos objetos da camada de acesso a dados, efetuando a inserção ou atualização dos dados no banco de dados. Se uma entidade já existir no banco de dados, os seus dados devem ser integrados e atualizados. A integração deve resultar na união dos atributos apresentados por diferentes sites. Deve ser mantido o histórico de dados relevantes para gerar gráficos de evolução de produtos, registrando-se a data de coleta (e.g., evolução de preços, de quantidade de itens vendidos, de avaliações etc.). Para dados duplicados, mas inconsistentes (i.e., atributos correspondentes a uma mesma entidade com dados diferentes), defina uma das fontes como a fonte "master" para o atributo.

O sistema deve também ter uma interface pública, com objetivo apresentar relatórios diversos a partir dos dados armazenados, com dados sumarizados,

tabelas e gráficos. Deve ser possível também fazer buscas de filtragem e ver os relatórios correspondentes aos resultados dos filtros.

### Processo de avaliação

O projeto pode ser feito em duplas. A avaliação será do trabalho e da atuação individual de cada membro do grupo, portanto, as notas dos membros de um mesmo grupo podem ser diferentes.

Serão realizadas duas entregas, que correspondem a a ciclos de desenvolvimento subsequentes:

a) 2º bimestre: interface de cadastro de websites, crawling, integração de dados e camada de acesso a dados e controllers correspondentes. Os dados de alguns crawlings devem já ter sido carregados.

b) 3º bimestre: interface pública e finalização do sistema

As entregas serão feitas por meio de apresentações agendadas com cada grupo. O processo de desenvolvimento será parte da nota, a ser verificado por commits no repositório de código do grupo, reuniões e pré-avaliações.