



Banco de dados - Avaliação 1

Descrição das fontes, da ferramenta de extração e
dos dados à serem trabalhados

Guilherme C. F. Mello

4 de juny de 2018

Índex

1	Introdução	1
2	Fontes de dados	2
2.1	Seleção das Fontes	2
2.2	Levantamento dos dados disponíveis	2
2.3	Definindo dados à serem estudados	3
3	Análise das informações	4
3.1	Ferramenta de extração de dados	4
3.2	Estatísticas à serem desenvolvidas	4
4	Bibliografia	5

1 Introdução

Neste trabalho serão estudados os dados referentes às mídias digitais de jogos comercializadas via web, para isso, será utilizada a ferramenta de extração de dados *open source* chamada *Scrapy*. Com os dados em mãos, serão feitos estudos à respeito de suas informações específicas em relação ao preço, data de lançamento e desenvolvedoras.

2 Fontes de dados

2.1 Seleção das Fontes

Para levantar as informações necessárias para o estudo, o ideal é carregar os dados de *e-commerces* de mídias digitais, pois em sua maioria, possuem muitas informações relevantes em relação aos jogos, seus desenvolvedores e compradores. Após um levantamento geral de fontes, foram encontradas as seguintes lojas virtuais: *Steam, Origin, Nuuvem, Greeman Gaming, GamersGate, Blizzard, Humble Bundle, Fanatical, Gog*.

Porém, não é possível fazer uma análise de qualidade sem selecionar criteriosamente as fontes com base nos seus dados fornecidos e credibilidade dos mesmos, portanto, foi montada a seguinte relação de dados mínimos que a loja virtual deveria disponibilizar: *Nome, preço, desenvolvedora, publicadora, data de lançamento, plataforma e gênero*. Levando em consideração a lista de dados mínimos e a estruturação das páginas chegamos em três fontes finais: *Steam, Nuuvem e Gamersgate*.

2.2 Levantamento dos dados disponíveis

Uma vez selecionadas as fontes, chegamos à seguinte tabela de dados que cada loja disponibiliza para nosso extrator:

Taulla 1: Dados disponibilizados por cada fonte

Steam	Nuuvem	Gamersgate
Nome	Nome	Nome
Preço	Preço	Preço
Quantidade de análises	Regiões para ativação da chave	Categorias
Porcentagem de análises boas	Ativação	DRM
Desenvolvedor	Data lançamento	Idiomas
Distribuidora	Desenvolvedor	Data lançamento
Marcadores populares	Distribuidora	Requisitos do sistema
Análises externas	Gênero	Desenvolvedora
Categorias	Modo de jogo (mult, single, coop)	Publicadora
Gêneros	Descrição geral	Plataforma
Data de lançamento	idiomas	Descrição geral
Descrição geral	Requisitos de sistema	
Prêmios do jogo	Classificação indicativa	
Requisitos do sistema	Plataforma	
Análise de usuários		
Plataformas suportadas		

2.3 Definindo dados à serem estudados

Com base na tabela "*Dados disponibilizados por cada fonte*", e utilizando a noção de interseção foi decidido extrair os seguintes dados dos três sites:

Taula 2: Informações por produto

Dados por jogos

Nome

Preço

Data de lançamento

Desenvolvedora

Publicadora

Categorias

Os estudos e extrações de dados serão focados nos itens mais vendidos de cada *e-commerce*, pois serão levantados dados mais interessantes.

3 Análise das informações

3.1 Ferramenta de extração de dados

Foi decidido utilizar a ferramenta de extração de dados *Scrapy*, pois, por ser *opensource*, tem muita informação fácil de encontrar e apoio da comunidade. Como foi definido pelo professor, serão gerados arquivos *Json* para organização e gestão das informações capturadas.

3.2 Estatísticas à serem desenvolvidas

Uma vez já definidas as informações de extração, serão geradas as seguintes estatísticas:

- Média de preço de jogo por Desenvolvedora
- Relação entre preço e data de lançamento
- Categorias de jogos mais vendidos
- Média geral de preço por site
- Relação preço e categorias

4 Bibliografia

<https://docs.scrapy.org/en/latest/>

<https://www.devmedia.com.br/json-tutorial/25275>