



02 de setembro de 2019

Tempo: 10 dias

Trabalho Prático I

---

### Especificação do Trabalho

1. A ParAmazon é uma grande empresa de comércio eletrônico do Gran-Pará. No seu site, as pessoas costumam postar comentários sobre diversos produtos. Na categoria de câmeras, há centenas de milhares de comentários. Porém, a empresa tem tido dificuldades em buscar (consultar) os aspectos mais frequentemente comentados pelos usuários devido ao grande volume de dados. Neste cenário, implemente um método que retorne os 10 aspectos mais frequentemente comentados pelos usuários.

O trabalho deverá ser realizado em **duplas** de alunos da disciplina de AED II. Segue abaixo a especificação do trabalho:

1. Fazer o download do dataset no link abaixo:

<https://we.tl/t-HCzQSXs1lP>

O arquivo possui 1.499.556 opiniões, onde cada linha representa uma opinião. Abaixo segue um exemplo de linha (opinião):

```
{4: (u'I've owned the Samsung for about a year and a half, and I just think the picture quality is awful especially shooting indoors or in lower light.', [(u'picture quality', u'awful', -1.0)])}
```

O valor 4 (quatro) indica o número do comentário. A sentença ‘‘I’ve owned the Samsung for about a year and a half, and I just think the picture quality is awful especially shooting indoors or in lower light.’’ representa o texto escrito pelo usuário. Os termos (picture quality’, u’awful’, -1.0) representam os aspectos, os sentimentos e a polaridade que foram extraídos da opinião.

Existem algumas sentenças que não têm opinião, tal como ocorre em {4: (‘‘Even playing around with the settings, trying higher pixels, etc.’’, [])}. Nesses casos, obviamente, não haverá aspectos. Além disso, em algumas sentenças será possível encontrar mais de uma opinião. Nesses casos, cada opinião terá um aspecto e que deverá ser considerado.

2. Implemente três algoritmos de ordenação para que possam ser exibidos os 10 aspectos mais frequentes do conjunto de dados. Considere os seguintes algoritmos: *quicksort*, *bucket sort* e *heap sort*. Faça uma comparação entre os algoritmos em relação ao tempo para execução e o número de trocas. Inclua nessa comparação o método de ordenação padrão disponível na linguagem de programação Python.

Reporte também uma variação da entrada de dados e veja se isso afeta o resultado final dos métodos considerados nesse trabalho. Além disso, alguns dos métodos de ordenação têm comportamento distinto de acordo com algumas configurações. Explore essas possíveis variações e reporte os resultados.

Você **não** deve usar implementações prontas. Faça a criação do seu próprio tipo abstrato de dados (TAD).

Recomendação: antes de propor as funções e os tamanhos, é importante conhecer melhor os dados que serão manipulados. Isso ajuda bastante para se ter boas ideias.

Os resultados obtidos deverão ser analisados e compreendidos pelos alunos.

3. Os resultados obtidos serão apresentados e discutidos em sala de aula. A apresentação será feita em dupla, conforme agendamento do professor. Deverá ser apresentado um gráfico (ou tabela) para facilitar o entendimento. Escolha unidades de medida apropriadas para o gráfico. Nessa apresentação, a dupla deverá explicar os resultados obtidos.

#### 4. Entrada e saída:

A correção (verificação) do programa será feita manualmente. Portanto, o programa deve ser desenvolvido para permitir que o professor possa executar os algoritmos listados anteriormente. Abaixo seguem exemplos de comandos esperados para interagir com o programa:

```
python -m 'algoritmo' main.py
```

A versão padrão do Python será a 2.7. Caso a implementação tenha sido em outra versão, a dupla deverá informar no relatório.

#### 5. Regras:

- Os trabalhos deverão ser realizados em dupla.
  - O início do trabalho será no dia **03 de setembro de 2019** (10:00 horas) e a entrega será no dia **10 de setembro de 2019** (até as 10 horas).
  - Deverá ser encaminhado para o e-mail `tmelo@uea.edu.br` do professor o relatório, os slides e o código-fonte. O relatório também deverá ser entregue impresso no dia 10 de setembro de 2019. Os três arquivos deverão estar compactados (.zip) e o título do e-mail deverá ser *“Trabalho Pratico I - AED 2”*.
  - Cada dupla terá **até 5 minutos** para apresentar o trabalho. A defesa deve ser preparada para esse tempo. Nesta defesa, a dupla deverá apresentar e justificar os resultados obtidos e as configurações utilizadas.
  - Esse trabalho será composto por duas notas: a implementação e a defesa. Estilo de programação e técnicas adotadas serão consideradas na avaliação da implementação. Segurança e conhecimento demonstrado na defesa serão considerados como critérios de avaliação.
  - A discussão sobre o trabalho deverá acontecer **somente** entre a dupla.
-