

## Relatório Atividade 2 - RII/2018-02

Aluno: Wesley Pereira da Silva

Professor: Elias Oliveira

T2) Tendo como exemplo a transformação do texto-documento em vetores feita no artigo (OLIVEIRA et al., 2007), utilize o indexador p-indexer para transformar os documentos da base de dados `orthogonals*.txt`, disponível em (<http://www.inf.ufes.br/~elias/dataSets/basic-datasets.tar.gz>). Esta atividade terá duas partes: A) Escreva um relatório de como utilizar esse utilitário para realizar essa atividade, desde a entrada na conta de usuário no LCAD, parâmetros necessários, arquivos gerados por conta da execução do p-indexer e suas respectivas importâncias para o processo de indexação, use também a referência (BÜTTCHER; CLARKE; CORMACK, 2010, Cap. 1 e 2) para fundamentar seus argumentos. B) Crie um *script* para gerar um gráfico com os pontos-documentos em um espaço  $\mathbb{R}^3$ . Submeta o relatório como um arquivo PDF, os *scripts* e o gráfico gerado pelo *script*.

1. Acessar o computador de acesso do lcad
  - a. `ssh rii-lcad@lcad.inf.ufes.br`
2. Acessar o computador que tem o aLine
  - a. `ssh rii-lcad@10.9.8.10`
3. Fazer download e descompactar o dataset
  - a. `wget http://www.inf.ufes.br/~elias/dataSets/basic-datasets.tar.gz`
  - b. `tar -xvzf basic-datasets.tar.gz`
4. Gerar a os arquivos de indice para o aLine
  - a. `find |grep txt > indexBasicDataset.dat`
5. Rodar o comando de indexação do aLine
  - a. `aLine -i -l indexBasicDataset.dat -d T2-Resultado`
6. Exportar as features do documentos
  - a. `aLine --convert -d T2-Resultado`
7. Rodar o programa em python que gera o gráfico.
  - a. `python T2-Wesley.py`

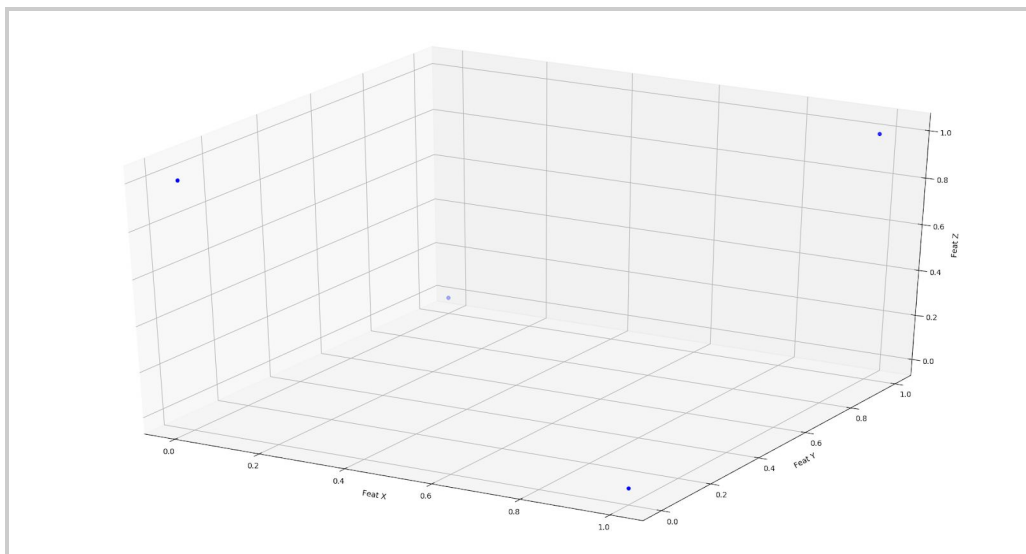


Figura 1 - Projeção dos pontos dos textos indexados