

Com o objetivo de guiar vocês na escrita do relatório, segue algumas documentações mínimas que devem estar nesse relatório:

- Explicação das classes principais bem como os métodos: Esse detalhamento é importante para que eu possa entender o código que foi criado. O código também deverá estar devidamente comentado e legível para facilitar o entendimento. Para facilitar o entendimento, uma sugestão é fazer um algoritmo em alto nível de como foi realizado o processamento/indexação e quando cada método no código foi chamado.
- Após a criação o índice, ele deve ser gravado em arquivo (binário ou texto, como preferir). Favor disponibilizar o link para que eu possa fazer o download do arquivo (via dropbox, por ex). Além de instruções completas para que eu possa abrí-lo.
- Discussão sobre a decisão de implementação adotada: Quais foram os principais desafios e soluções? Qual é a vantagem/desvantagem das suas soluções sob as outras alternativas (por exemplo, uso do índice em memória principal x ocorrência de termos em memória secundária)? O que você melhoraria no seu código para diminuir o consumo de memória ou deixá-lo mais eficiente?
- Ainda sobre decisão de implementação: se tivéssemos milhões de documentos e a lista de ocorrência não coubesse em memória principal, como você mudaria o processo criação e a estrutura deste índice? Descreva detalhadamente.
- Quais são as bibliotecas externas utilizadas? Explique o funcionamento da técnica de stemming adotada.
- Qual foi a estrutura do índice utilizado? Quanto MB de ram sua solução gastou? Em quanto tempo foi realizado a indexação? Qual foi a média por documento?

Caso tenham alguma dúvida, mandem email. Favor esclarecer dúvidas com uma certa antecedência, pois o prazo não será alterado por causa de dúvidas de última hora no relatório ou na implementação.