

TCC 1

● Utilização de arquivos

Um mecanismo que estou implementando é a utilização de arquivos para não se perder os links que já foram rastreados, pois até o momento, todos os links estão ficando em cache e assim que termina a execução do programa, todos são perdidos e temos que coletar todos de novo.

Entretanto com a utilização de arquivos não teremos mais que começar do zero, já aproveitando links que foram rastreados. Um problema sobre a utilização de arquivos são páginas que foram alteradas entre a execução do programa, elas não serão novamente rastreadas. Entretanto com este mecanismo poderemos chegar até o final do processo sem ter que manter o computador ligado sem parar por horas.

● Outros problemas

- A revista coringa tem páginas que possuem comentários e essas páginas são extremamente grandes e pesadas, demorando muito o processamento delas.
Ex:
https://www.jornalismo.ufop.br/revistacuringa/?attachment_id=19#comment-39141
- Eventos da UFOP geram vários e vários links no calendário
<https://ppgmqmg.ufop.br/calendar/week/2023-02?type=month&3Bamp%3Bmonth=2023-02&3Bamp%3Bmini=2023-03&3Bmonth=2023-02&month=2023-02>
- Novamente o calendário, para quase todos os subdomínios da UFOP, vão ter calendários com os eventos, trazendo eventos de vários anos atrás.
<https://ppgmqmg.ufop.br/calendar/week/2023-02?type=month&3Bamp%3Bmonth=2023-02&3Bamp%3Bmini=2023-03&3Bmonth=2023-02&month=2023-02>
-