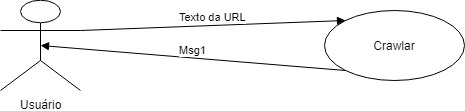
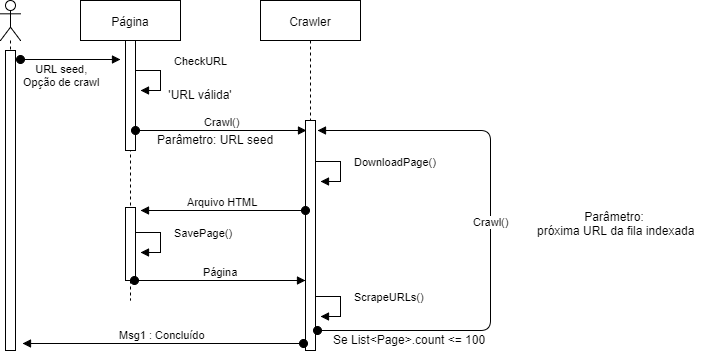
1. **História**

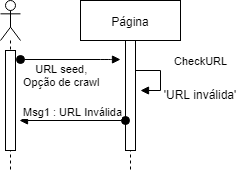
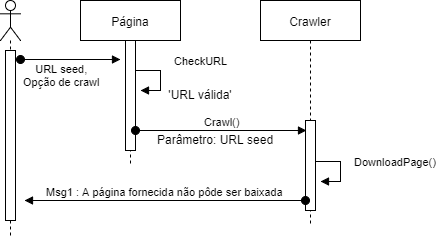
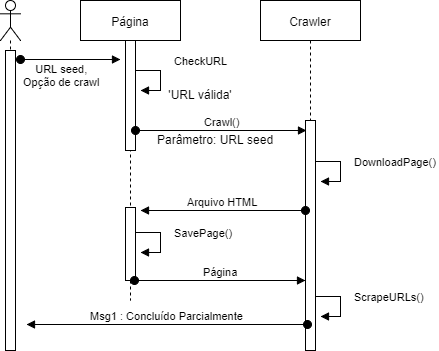
O sistema de crawler deve receber o texto de uma URL inicial do usuário. Com essa URL inicial, o sistema deve pesquisar e arquivar essa página e outras páginas conectadas direta ou indiretamente a ela, até atingir um máximo de 100 páginas visitadas, salvando as páginas em arquivos HTML em uma pasta nomeada com o horário em que o processo foi terminado.

1. **Identificação dos substantivos**
   1. Sistema de crawler
   2. Texto
   3. URL
   4. Usuário
   5. Página
   6. Ela
   7. Máximo
   8. Arquivos HTML
   9. Pasta
   10. Horário
   11. Processo
2. **Identificação dos substantivos relevantes**
   1. Texto
   2. URL
   3. Usuário
   4. Página
   5. Arquivos HTML
3. **Identificar Atores**
   1. Usuário
4. **Identificar objetos**
   1. URL
   2. Página
5. **Identificar propriedades**
   1. Texto [da URL]
   2. Arquivo HTML
6. **Relacionamento de objetos e propriedades**
   1. URL
      1. Texto
   2. Página
      1. Arquivo HTML
7. **Identificar verbos de ação**
   1. Receber uma URL inicial do usuário
   2. Pesquisar e arquivar essa página e outras páginas conectadas direta ou indiretamente a ela, até atingir um máximo de 100 páginas visitadas, salvando as páginas em arquivos HTML em uma pasta nomeada com o horário em que o processo foi terminado.
8. **Tabela de Eventos (Arquivo Excel também no git)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Ator** | **Entrada** | **Saída** | **Ação** |
| **Usuário** | **Texto da URL Inicial** | **Mensagem 1** | **CrawlAndSave()** |

1. **Casos de uso para o usuário**
   1. Crawlar
2. **Descrever fluxo normal e alternativos dos casos de uso** (Apenas um nesse caso)
   1. Fluxo Normal
      1. O usuário fornece o texto da URL seed inicial
      2. O sistema checa se a URL fornecida é válida
      3. O sistema acessa essa URL e baixa o arquivo HTML da página
      4. O sistema salva essa página na pasta desse crawl
      5. O sistema reconhece outras URLs dentro do arquivo HTML dessa página e as salva na lista de URL “fronteira do crawl” de forma indexada usando uma lógica opcional de largura ou profundidade
      6. O sistema repete o processo de 2 a 4 até que o número de páginas já arquivadas atinja 100
      7. O sistema informa o usuário que o processo foi concluído



* 1. Fluxo alternativo: Caso da URL fornecida inválida (partindo do passo 2)
     1. O sistema retorna uma mensagem de URL inválida por validação própria ao usuário
     2. Finalizar caso de uso
  2. Fluxo alternativo: Problemas de acesso à página da URL fornecida (partindo do passo 3)
     1. O sistema retorna a mensagem de URL inválida por erro de acesso ao usuário
     2. Finalizar caso de uso
  3. Fluxo alternativo: Não há nenhuma outra URL na página HTML sendo reconhecida (passo 5)
     1. O sistema retorna a mensagem de fim do crawl para o usuário
     2. Finalizar caso de uso

1. **Diagrama de Classes**

