As configurações do Scrapy são usadas para personalizar o comportamento de todos os componentes do Scrapy, incluindo o núcleo, extensões, pipelines e spiders. A infraestrutura das configurações fornece um namespace global de mapeamentos de chave-valor que o código pode usar para obter valores de configuração.

As configurações do Scrapy são definidas em um arquivo de configuração, que geralmente é chamado de settings.py. Este arquivo deve ser colocado no diretório raiz do seu projeto Scrapy.

Aqui está uma lista das configurações do Scrapy mais comuns, com uma explicação de cada uma delas:

* BOT\_NAME (nome do bot): O nome do seu bot Scrapy. Isso é usado para identificar seu bot nos logs e outros arquivos.
* CLOSESPIDER\_ERRORCOUNT (contagem de erros de fechamento): O número de erros que Scrapy permitirá antes de fechar o spider.
* CLOSESPIDER\_ITEMCOUNT (contagem de itens de fechamento): O número de itens que Scrapy permitirá que o spider colete antes de fechá-lo.
* CLOSESPIDER\_PAGECOUNT (contagem de páginas de fechamento): O número de páginas que Scrapy permitirá que o spider explore antes de fechá-lo.
* CLOSESPIDER\_TIMEOUT (tempo limite de fechamento): O número de segundos que Scrapy esperará pelo spider terminar de coletar dados antes de fechá-lo.
* CONCURRENT\_ITEMS (itens simultâneos): O número máximo de itens que Scrapy processará simultaneamente por resposta.
* CONCURRENT\_REQUESTS (solicitações simultâneas): O número máximo de solicitações que Scrapy fará simultaneamente.
* CONCURRENT\_REQUESTS\_PER\_DOMAIN (solicitações simultâneas por domínio): O número máximo de solicitações que Scrapy fará simultaneamente para um único domínio.
* CONCURRENT\_REQUESTS\_PER\_IP (solicitações simultâneas por IP): O número máximo de solicitações que Scrapy fará simultaneamente para um único endereço IP.
* DEPTH\_LIMIT (limite de profundidade): A profundidade máxima de links que Scrapy seguirá.
* DEPTH\_PRIORITY (prioridade de profundidade): Uma prioridade para links com profundidades mais baixas.
* DEPTH\_STATS\_VERBOSE (estatísticas de profundidade detalhadas): Habilita estatísticas detalhadas sobre a profundidade dos links explorados.
* DOWNLOAD\_DELAY (atraso de download): Um atraso em segundos entre cada solicitação.
* DOWNLOAD\_MAXSIZE (tamanho máximo de download): O tamanho máximo de um download, em bytes.
* DOWNLOAD\_TIMEOUT (tempo limite de download): O número de segundos que Scrapy esperará por uma resposta antes de desistir.
* DUPEFILTER\_CLASS (classe de filtro de duplicados): A classe de filtro de duplicados a ser usada.
* HTTPPROXY\_ENABLED (proxy HTTP habilitado): Habilita o uso de um proxy HTTP.
* ITEM\_PIPELINES (pipelines de itens): Uma lista de pipelines de itens a serem usados.
* LOG\_ENABLED (registro habilitado): Habilita o registro.
* LOG\_LEVEL (nível de registro): O nível de registro, como DEBUG, INFO, WARNING, ERROR ou CRITICAL.
* MEMUSAGE\_ENABLED (uso de memória habilitado): Habilita a coleta de estatísticas de uso de memória.
* MEMUSAGE\_LIMIT\_MB (limite de uso de memória em MB): O limite de uso de memória em megabytes. Se o uso de memória exceder este limite, Scrapy encerrará o spider.
* MEMUSAGE\_WARNING\_MB (aviso de uso de memória em MB): O valor de uso de memória em megabytes em que Scrapy emitirá um aviso.
* REDIRECT\_ENABLED (redirecionamento habilitado): Habilita o seguimento de redirecionamentos.
* RETRY\_ENABLED (repetição habilitada): Habilita a repetição de solicitações com falha.
* ROBOTSTXT\_OBEY (obedecer ao robots.txt): Habilita a obediência ao arquivo robots.txt.
* SCHEDULER (agendador): A classe de agendador a ser usada.
* SPIDER\_MIDDLEWARES (middlewares de spider): Uma lista de middlewares de spider a serem usados.
* STATS\_CLASS (classe de estatísticas): A classe de coletor de estatísticas a ser usada.
* STATS\_DUMP (despejo de estatísticas): Habilita o despejo de estatísticas para um arquivo.
* USER\_AGENT (agente de usuário):