“””

Configurações do Scrapy com uma breve explicação de cada uma delas:

“””

ADDONS: Configurações adicionais, usadas para adicionar funcionalidades extras ao Scrapy.

AJAXCRAWL\_ENABLED: Define se a renderização de páginas AJAX (usada para carregar conteúdo dinâmico) está ativada.

ASYNCIO\_EVENT\_LOOP: Permite configurar o loop de eventos asyncio personalizado se estiver usando asyncio.

AUTOTHROTTLE\_DEBUG: Ativar a depuração do autotrottle, que ajusta automaticamente a taxa de download.

AUTOTHROTTLE\_ENABLED: Ativar o autotrottle, que ajusta automaticamente a taxa de download.

AUTOTHROTTLE\_MAX\_DELAY: O atraso máximo que o autotrottle pode adicionar entre as solicitações.

AUTOTHROTTLE\_START\_DELAY: O atraso inicial aplicado pelo autotrottle.

AUTOTHROTTLE\_TARGET\_CONCURRENCY: A taxa de solicitações-alvo usada pelo autotrottle.

BOT\_NAME: O nome do bot Scrapy.

CLOSESPIDER\_ERRORCOUNT: Fecha a aranha se um número de erros for atingido.

CLOSESPIDER\_ITEMCOUNT: Fecha a aranha após encontrar um número específico de itens.

CLOSESPIDER\_PAGECOUNT: Fecha a aranha após visitar um número específico de páginas.

CLOSESPIDER\_TIMEOUT: Fecha a aranha após um período de tempo específico.

COMMANDS\_MODULE: Módulo usado para registrar comandos personalizados.

COMPRESSION\_ENABLED: Ativa a compressão de saída.

CONCURRENT\_ITEMS: O número máximo de itens a serem processados em paralelo.

CONCURRENT\_REQUESTS: O número máximo de solicitações a serem processadas em paralelo.

CONCURRENT\_REQUESTS\_PER\_DOMAIN: O número máximo de solicitações simultâneas permitidas por domínio.

CONCURRENT\_REQUESTS\_PER\_IP: O número máximo de solicitações simultâneas permitidas por IP.

COOKIES\_DEBUG: Ativa a depuração de cookies.

COOKIES\_ENABLED: Ativa o suporte a cookies.

DEFAULT\_ITEM\_CLASS: A classe padrão para itens se não especificada em aranhas.

DEFAULT\_REQUEST\_HEADERS: Os cabeçalhos HTTP padrão enviados em todas as solicitações.

DEPTH\_LIMIT: A profundidade máxima de busca em páginas.

DEPTH\_PRIORITY: Prioridade atribuída às solicitações com base em sua profundidade.

DEPTH\_STATS\_VERBOSE: Ativa estatísticas detalhadas sobre a profundidade das solicitações.

DNSCACHE\_ENABLED: Ativa o cache DNS.

DNSCACHE\_SIZE: Tamanho máximo do cache DNS.

DNS\_RESOLVER: A classe usada para resolver nomes de domínio.

DNS\_TIMEOUT: O tempo limite de resolução DNS.

DOWNLOADER: A classe do downloader usada para baixar conteúdo da web.

DOWNLOADER\_CLIENTCONTEXTFACTORY: A classe usada para criar contextos SSL.

DOWNLOADER\_CLIENT\_TLS\_CIPHERS: Configura as cifras SSL usadas.

DOWNLOADER\_CLIENT\_TLS\_METHOD: Configura o método SSL usado.

DOWNLOADER\_CLIENT\_TLS\_VERBOSE\_LOGGING: Ativa logs detalhados para TLS.

DOWNLOADER\_HTTPCLIENTFACTORY: A classe usada para criar clientes HTTP.

DOWNLOADER\_MIDDLEWARES: Middleware usado no processo de download.

DOWNLOADER\_STATS: Ativa estatísticas do downloader.

DOWNLOAD\_DELAY: O atraso entre as solicitações.

DOWNLOAD\_FAIL\_ON\_DATALOSS: Define se a aranha falhará ao encontrar perda de dados durante o download.

DOWNLOAD\_HANDLERS: Handlers usados para diferentes tipos de respostas.

DOWNLOAD\_MAXSIZE: Tamanho máximo dos downloads.

DOWNLOAD\_TIMEOUT: Tempo limite para downloads.

DOWNLOAD\_WARNSIZE: Tamanho máximo dos downloads para acionar avisos.

DUPEFILTER\_CLASS: A classe usada para filtrar URLs duplicadas.

EDITOR: O editor usado para abrir aranhas.

EXTENSIONS: Extensões Scrapy que estão ativadas.

EXTENSIONS\_BASE: Extensões base que estão ativadas.

FEEDS: Configurações para exportar dados em feeds (por exemplo, JSON, CSV).

FEED\_EXPORTERS: Exportadores de feeds ativados.

FEED\_EXPORTERS\_BASE: Exportadores de feeds base.

FEED\_EXPORT\_BATCH\_ITEM\_COUNT: Número máximo de itens por lote exportado.

FEED\_EXPORT\_ENCODING: Codificação usada para exportação de feeds.

FEED\_EXPORT\_FIELDS: Campos a serem exportados.

FEED\_EXPORT\_INDENT: Indentação usada na exportação de feeds.

FEED\_STORAGES: Armazenamento de feeds ativado.

FEED\_STORAGES\_BASE: Armazenamento de feeds base.

FEED\_STORAGE\_FTP\_ACTIVE: Ativa o modo FTP ativo para armazenamento de feeds.

FEED\_STORAGE\_GCS\_ACL: Política de controle de acesso para armazenamento de feeds no Google Cloud Storage.

FEED\_STORAGE\_S3\_ACL: Política de controle de acesso para armazenamento de feeds no Amazon S3.

FEED\_STORE\_EMPTY: Armazena feeds vazios.

FEED\_TEMPDIR: Diretório temporário para armazenamento de feeds.

FEED\_URI\_PARAMS: Parâmetros adicionais para feeds.

FILES\_STORE\_GCS\_ACL: Política de controle de acesso para armazenamento de arquivos no Google Cloud Storage.

FILES\_STORE\_S3\_ACL: Política de controle de acesso para armazenamento de arquivos no Amazon S3.

FTP\_PASSIVE\_MODE: Ativa o modo FTP passivo.

FTP\_PASSWORD: Senha usada para autenticação FTP.

FTP\_USER: Nome de usuário usado para autenticação FTP.

GCS\_PROJECT\_ID: ID do projeto Google Cloud Storage.

HTTPCACHE\_ALWAYS\_STORE: Armazena todas as respostas em cache, mesmo que elas não tenham cabeçalhos de cache.

HTTPCACHE\_DBM\_MODULE: Módulo de DBM usado para armazenar em cache.

HTTPCACHE\_DIR: Diretório de armazenamento em cache.

HTTPCACHE\_ENABLED: Ativa o armazenamento em cache de respostas HTTP.

HTTPCACHE\_EXPIRATION\_SECS: Tempo de expiração das respostas em cache.

HTTPCACHE\_GZIP: Ativa a compactação GZIP para respostas em cache.

HTTPCACHE\_IGNORE\_HTTP\_CODES: Códigos de status HTTP para ignorar no armazenamento em cache.

HTTPCACHE\_IGNORE\_MISSING: Ignora respostas ausentes no armazenamento em cache.

HTTPCACHE\_IGNORE\_RESPONSE\_CACHE\_CONTROLS: Ignora os cabeçalhos de controle de cache nas respostas.

HTTPCACHE\_IGNORE\_SCHEMES: Esquemas de URL para ignorar no armazenamento em cache.

HTTPCACHE\_POLICY: Política de armazenamento em cache a ser usada.

HTTPCACHE\_STORAGE: Armazenamento de cache a ser usado.

HTTPPROXY\_AUTH\_ENCODING: Codificação usada para autenticação proxy HTTP.

HTTPPROXY\_ENABLED: Ativa o suporte a proxies HTTP.

IMAGES\_STORE\_GCS\_ACL: Política de controle de acesso para armazenamento de imagens no Google Cloud Storage.

IMAGES\_STORE\_S3\_ACL: Política de controle de acesso para armazenamento de imagens no Amazon S3.

ITEM\_PIPELINES: Pipelines a serem aplicados aos itens após a raspagem.

ITEM\_PROCESSOR: Gerenciador de pipelines de itens.

LOGSTATS\_INTERVAL: Intervalo de tempo entre os registros de estatísticas no log.

LOG\_DATEFORMAT: Formato da data e hora nos registros do log.

LOG\_ENABLED: Ativa ou desativa o registro no log.

LOG\_ENCODING: Codificação usada para registros de log.

LOG\_FILE: Arquivo onde os registros do log são armazenados.

LOG\_FILE\_APPEND: Define se os registros de log são anexados a um arquivo existente.

LOG\_FORMAT: Formato dos registros de log.

LOG\_FORMATTER: Formatador de registros de log.

LOG\_LEVEL: Nível de detalhe dos registros de log.

LOG\_SHORT\_NAMES: Ativa nomes curtos nos registros de log.

LOG\_STDOUT: Ativa o registro no console.

MAIL\_FROM: Endereço de e-mail de origem para notificações de e-mail.

MAIL\_HOST: Host do servidor de e-mail para notificações de e-mail.

MAIL\_PASS: Senha do servidor de e-mail para notificações de e-mail.

MAIL\_PORT: Porta do servidor de e-mail para notificações de e-mail.

MAIL\_USER: Nome de usuário do servidor de e-mail para notificações de e-mail.

MEMDEBUG\_ENABLED: Ativa a depuração de uso de memória.

MEMDEBUG\_NOTIFY: Lista de destinatários para notificações de uso de memória.

MEMUSAGE\_CHECK\_INTERVAL\_SECONDS: Intervalo de tempo entre verificações de uso de memória.

MEMUSAGE\_ENABLED: Ativa o monitoramento de uso de memória.

MEMUSAGE\_LIMIT\_MB: Limite máximo de uso de memória.

MEMUSAGE\_NOTIFY\_MAIL: Lista de destinatários para notificações de uso de memória por e-mail.

MEMUSAGE\_WARNING\_MB: Limite de aviso de uso de memória.

METAREFRESH\_ENABLED: Ativa o redirecionamento de metaatualização.

METAREFRESH\_IGNORE\_TAGS: Tags HTML que são ignoradas pelo redirecionamento de metaatualização.

METAREFRESH\_MAXDELAY: Atraso máximo entre redirecionamentos de metaatualização.

NEWSPIDER\_MODULE: Módulo para criar novas aranhas.

RANDOMIZE\_DOWNLOAD\_DELAY: Aleatoriza o atraso entre as solicitações.

REACTOR\_THREADPOOL\_MAXSIZE: Tamanho máximo da threadpool do Twisted Reactor.

REDIRECT\_ENABLED: Ativa redirecionamentos de URL.

REDIRECT\_MAX\_TIMES: Número máximo de redirecionamentos permitidos.

REDIRECT\_PRIORITY\_ADJUST: Ajuste de prioridade para solicitações de redirecionamento.

REFERER\_ENABLED: Ativa a inclusão do cabeçalho Referer nas solicitações.

REFERRER\_POLICY: Política de cabeçalho Referer usada.

REQUEST\_FINGERPRINTER\_CLASS: Classe usada para calcular impressões digitais de solicitação.

REQUEST\_FINGERPRINTER\_IMPLEMENTATION: Versão da implementação do cálculo de impressão digital de solicitação.

RETRY\_ENABLED: Ativa a reexecução de solicitações em caso de erros.

RETRY\_EXCEPTIONS: Exceções que acionam tentativas de reexecução.

RETRY\_HTTP\_CODES: Códigos de status HTTP que acionam tentativas de reexecução.

RETRY\_PRIORITY\_ADJUST: Ajuste de prioridade para solicitações de reexecução.

RETRY\_TIMES: Número máximo de tentativas de reexecução.

ROBOTSTXT\_OBEY: Ativa a obediência ao robots.txt.

ROBOTSTXT\_PARSER: Parser usado para interpretar o arquivo robots.txt.

ROBOTSTXT\_USER\_AGENT: User-agent usado ao acessar o robots.txt.

SCHEDULER: Classe do escalonador usado.

SCHEDULER\_DEBUG: Ativa a depuração do escalonador.

SCHEDULER\_DISK\_QUEUE: Classe da fila de disco usada no escalonador.

SCHEDULER\_MEMORY\_QUEUE: Classe da fila de memória usada no escalonador.

SCHEDULER\_PRIORITY\_QUEUE: Classe da fila de prioridade usada no escalonador.

SCRAPER\_SLOT\_MAX\_ACTIVE\_SIZE: Tamanho máximo de slots de raspagem ativos.

SPIDER\_CONTRACTS: Contratos usados nas aranhas.

SPIDER\_CONTRACTS\_BASE: Contratos de aranhas base.

SPIDER\_LOADER\_CLASS: Classe usada para carregar aranhas.

SPIDER\_LOADER\_WARN\_ONLY: Define se a ausência de aranhas é apenas um aviso.

SPIDER\_MIDDLEWARES: Middleware usado nas aranhas.

SPIDER\_MIDDLEWARES\_BASE: Middleware base usado nas aranhas.

SPIDER\_MODULES: Módulos que contêm aranhas.

STATSMAILER\_RCPTS: Destinatários para estatísticas de email.

STATS\_CLASS: Classe usada para coletar estatísticas.

STATS\_DUMP: Ativa o despejo de estatísticas.

TELNETCONSOLE\_ENABLED: Ativa o console Telnet.

TELNETCONSOLE\_HOST: Host do console Telnet.

TELNETCONSOLE\_PASSWORD: Senha do console Telnet.

TELNETCONSOLE\_PORT: Porta do console Telnet.

TELNETCONSOLE\_USERNAME: Nome de usuário do console Telnet.

TEMPLATES\_DIR: Diretório de modelos Scrapy.

TWISTED\_REACTOR: O Reactor Twisted a ser usado.

URLLENGTH\_LIMIT: Limite de comprimento de URL.

USER\_AGENT: O User-Agent padrão usado nas solicitações do Scrapy.