Algorithmus Innovation-Tool

1. Erhalte Main-URLs, die durchsucht werden sollen
2. MAIN-URL in \_websites\_searched aufnehmen
3. Suche für jede Main-URL die Sub-URLs bei der host übereinstimmt
   1. Sub-URL in Tabelle \_websites\_searched schreiben
   2. Suche für jede Sub-URL die Sub-Sub-URLs bei der host von Main-URL übereinstimmt
      1. Sub-Sub-URL in Tabelle \_websites\_searched schreiben
      2. Schleife beenden
   3. Schleife beenden
4. Alle zu durchsuchenden URLs aus Tabelle \_websites\_searched auslesen
5. Für jede URL:
   1. HTML-file erhalten
   2. Dieses Filtern (Sonderzeichen (außer „-„) und Mehrfachleerzeichen weglöschen)
   3. Text in Array umwandeln
   4. Array filtern (Stop-List, Länge = 1 oder 2, redundante Wörter)
   5. Für jedes Wort im Array
      1. Erhalte zugehörige Tabelle (Anfangsbuchstabe des Wortes)
      2. Abfrage mit „match-against“, ob es in Blacklist vorhanden ist
      3. Wenn in Blacklist vorhanden -> aus Array löschen
      4. Bei alle nicht gefunden Wörter:
         1. URL hinzufügen
         2. In Tabelle innovation\_check aufnehmen
6. Ausgabe der innovation\_check Tabelle
7. Nutzerinteraktion („checked“ = als Innovation markiert)
   1. „gecheckten“ Wörter
      1. in \_innovation\_found aufnehmen
      2. Zugehörige URL + id vom Wort zu \_innovation\_found\_urls hinzufügen
   2. „ungecheckte“ Wörter zu Blacklist (Tabellen a-z) hinzufügen