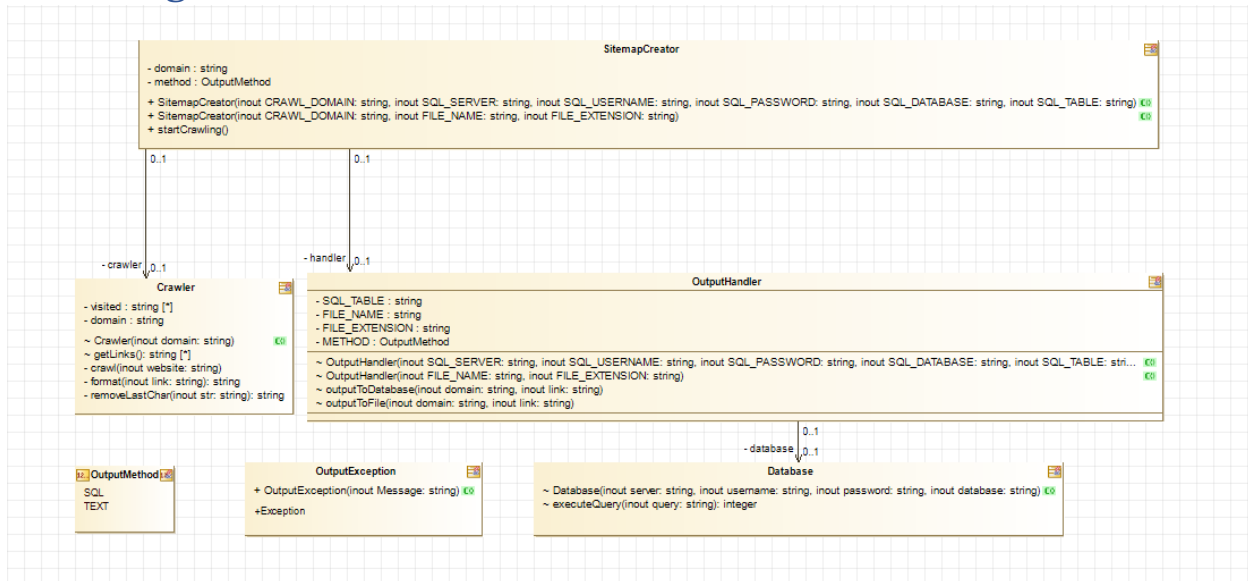


SitemapCreator

Uppgift

Uppgiften var att skapa ett valfritt projekt inom nätverk och/eller databaser. Utifrån de specifikationerna skapades ett projekt som hade i uppgift att inhämta alla åtkomliga sidor från en valfri hemsida. Detta skulle sedan exporteras antingen till en databas eller i textform till en fil.

Klassdiagram



JDOCS

Inkluderat i Github.

<https://filipolofsson.github.io/Sitemap-Creator/package-summary.html>

Diskussion

Funktion

Programmet börjar med att användaren väljer att initialisera ett objekt av SitemapCreator med en av de två konstruktorerna. En för varje exportmetod. Användaren kallar sedan startCrawling() på objektet av SitemapCreator med en domän. Detta kallar i sin tur Crawler.getLinks som kallar Crawler.crawl rekursivt tills ingen länk hittas på sidan som crawlas. Crawler.crawl använder sig av Crawler.format för att vara säker att alla länkar som hittas är konsekventa. Crawler.crawl sparar alla länkar som den besöker till en lista som returneras av Crawler.getLinks när rekursionen har stoppats. Efter SitemapCreatorklassen har listan över alla länkar som hittades så exporteras det antingen till en textfil med valbart namn och filtyp eller till en databas.

Problem

Det största problemet med applikationen var att inte alla länkar som hittas på sidor är likadana. T.ex. så kan en sida skriva sina länkar som *domän.se/index.php* medans en annan sida skriver deras länkar som */index.php*. Detta löstes genom att formatera länkarna så de ser likadana ut.

Buggar

Applikationen klarar i nuläget ej av att börja på i en mapp på en domän. Detta anses inte vara ett problem eftersom att applikationen är skapad för att kunna skapa en såkallad "sitemap", dessa startar alltid i document root.

Avslutning

Till nästa gång finns det en del saker som kan förbättras. Ett av de största förbättringsområdena är formateringen av länkar. Den metod som används i nuläget är långsam och dåligt gjort. Ett förslag skulle kunna vara att använda sig av regex för att implementera regler över vad som returneras av metoden.

Förutom detta så finns det möjlighet att i framtiden lägga till fler exporteringsalternativ, t.ex. FTP.

Länkar

Github - <https://github.com/FilipOlofsson/Sitemap-Creator>

Docs - <https://filipolofsson.github.io/Sitemap-Creator/>