**Lidl.co.uk store based promotion data scraping**

Issues:

Site’s store locator don’t generate individual store url, for a postcode search it displays 10 nearest stores but there is no absolute url available for those links.

Example of Url available for each store:

<a href="#">Select store to see offers</a>

Once one of the link is clicked it takes to that store page, site adds stores information in cookie and each time a category is selected it sends that cookie information along with that category link.

**Challenges:**

1. Bot don’t support to update cookie for each url, for example lidl.co.uk have 612 stores that means need to generate 612 cookies and need to update cookie for each store url.
2. Need to build urls and cookies for each store
3. Need to verify that the custom urls and cookies (Urls and cookies are built from store scraped information). The problem is returned html don’t have any information that confirms our requested url returned the exact store’s html. When we use browser and select one store on browser it displays the store address (My Lidl) but this store address information don’t available in the source code.

Example:

Url: <http://www.lidl.co.uk/en/index.htm?ar=2&changestore=1>

Url parameters:

Ar = ar

ChangeStoe = 1

Cookie: ar=ar=2;ar\_zip=London;ar\_city=NW 2 6ND;ar\_street=Edgware Road 408;nf=true

Cookie Parameters:

Ar = ar is available for each store scraping

Zip = City name

City = Zipcode

Ar\_street = Address field of scraping data

Nr = nr (true or false) when empty update as false

Referer: http://www.lidl.fr/fr/index.htm?country=FR&zipcode=75011&city=PARIS&street=130%20Rue%20du%20Chemin%20Vert&ar=10&nf=true&changestore=1

Custom Url:

<http://www.lidl.co.uk/en/index.htm?ar=Update_ar&changestore=1>

Cookie: ar=ar=2;ar\_zip=London;ar\_city=NW 2 6ND;ar\_street=Edgware Road 408;nf=true

Custom Cookie:

ar=ar=Update\_ar;ar\_zip=Update\_area;ar\_city=Update\_zipcode;ar\_street=Update\_address;nf=Update\_nf

Final String to updata:

insert into @urls Values('ar=ar=Update\_ar;ar\_zip=Update\_area;ar\_city=Update\_zipcode;ar\_street=Update\_address;nf=Update\_nf¬http://www.lidl.co.uk/en/index.htm?ar=Update\_ar&changestore=1',0)

insert into @urls Values('ar=ar=7;ar\_zip= Nelson;ar\_city=Update\_zipcode;ar\_street=Update\_address;nf=Update\_nf¬http://www.lidl.co.uk/en/index.htm?ar=Update\_ar&changestore=1',0)

**Conclusion :** It is possible to scrape store based promotion sites but need to decide how to update cookies for each url (need core changes in ScrapeingHerd and WebScraper). Also need to verify the returned html to make sure bot is scraping the store it supposed to.

Proposal :

\*\*\* Need to get cookie for each url

-- We can add cookie for each url in Url name field and from that field we can update cookie for each url.

Or we can add another field in ScraperSiteUrl table like UrlSettings which will contain Json Key Value Pairs and we can convert this Json using JObject and Use it the way we want.