**Vježba 1 – Crawler**

Realizirati aplikaciju za otkrivanje i popisivanje sadržaja povezanih linkovima (pojednostavljeni WebSpider ili WebCrawler). Aplikacija će kretati od zadane adrese (<https://faculty.cs.byu.edu/~rodham/cs240/crawler/> ili probajte stranicu iz primjera sa predavanja *www.optimazadar.hr*), primiti odgovor (u obliku HTTP odgovora sa HTML stranicom) i popisati sve linkove obradom odgovora. Nakon toga za svaki popisani link ponoviti postupak slanja zahtjeva na istu adresu sa različitom putanjom.

Stranica će se tretirati kao običan string gdje će se linkovi prepoznati po imenu atributa „href=“. Svaki link je ograničen navodnicima (jednostrukim ili dvostrukim). Svaki link treba dodati na listu „links“. Primjeri linkova u HTML kôdu:

<a href=“http:// www.optimazadar.hr/1280/djelatnost1280.html“> Home </a>

<a href=“http:// www.optimazadar.hr/1280/galerija1280.html“> Galerija </a>

<a href=“http:// www.optimazadar.hr/1280/reference1280.html“> Reference </a>

Za posjećivanje stranice preko socketa će se slati HTTP poruka (zahtjev) tipa:

**GET** *putanja* **HTTP/1.1**

**Host:** [*www.optimazadar.com*](http://www.optimazadar.com/) *,* [*www.faculty.cs.byu.edu*](https://faculty.cs.byu.edu/~rodham/cs240/crawler/)*)*

Ograničiti broj posjećenih linkova na 50 i sve linkove na kraju treba ispisati na konzoli. Ovisno o implementaciji, svi linkovi se neće uspjeti dohvatiti (provjeriti da li je status odgovora „200 OK“), takve linkove se može ignorirati.

Da bi se izbjeglo kruženje, prije dodavanja linka u listu provjerite da li se link već nalazi u listi.