|  |
| --- |
|  |
| Web Crawler |
|  |
| SOFTWARE DESIGN AND TEST |

**HANSUNG UNIV.**

2018 June 20

저자: 1291036 이정욱

Web Crawler

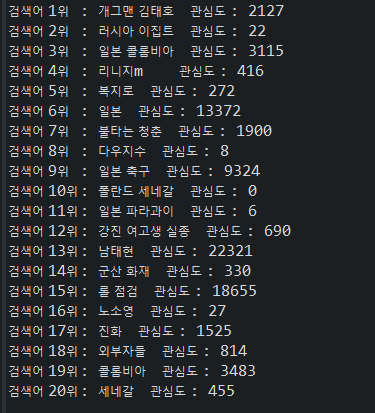
SOFTWARE DESIGN AND TEST

기능 설명

네이버 실시간 검색 순위를 가져와서 그걸 페이스북에 검색해서 그 키워드가 얼마나 관심을 받고 있는지, 입소문을 타고 있는지에 대한 정보를 수집한다.

동작 과정 설명

우선 네이버 데이터 랩 사이트에서 실시간 검색 결과를 찾아서 배열에 저장한다. 그리고 페이스북 웹페이지에 접속하여 그 키워드를 1위부터 20위까지 검색창에 검색한다. 그리고 검색 결과에서 각 게시물의 좋아요 수, 댓글 수, 공유 수를 가져와서 점수를 매긴다. 최종적으로 각 검색어의 관심도를 출력한다.

중요 코드 설명

driver.get("https://datalab.naver.com/keyword/realtimeList.naver?where=main");

for (i = 0; i < 20; i++) {

words[i] = new SearchWord();

words[i].setRank(i + 1);

words[i].setWord(driver.findElement(By.xpath("//div[@id='content']/div/div[3]/div/div/div[4]/div/div/ul/li[" + Integer.toString(i + 1) + "]/a/span")).getText());

}

* 네이버 데이터 랩 사이트에서 검색어들을 가져온다.

for (i = 0; i < 20; i++) {

(new WebDriverWait(driver, 30))

.until(ExpectedConditions.elementToBeClickable(driver.findElement(By.name("q"))));

while (true) {

try {

driver.findElement(By.name("q")).click();

break;

} catch (Exception e) {

// TODO: handle exception

}

}

interestPerWord = 0;

// System.out.println(words[i].getWord());

driver.findElement(By.name("q")).sendKeys(words[i].getWord());

driver.findElement(By.name("q")).sendKeys(Keys.ENTER);

* 페이스북 페이지에 가져온 검색어들을 검색한다.

List<WebElement> interestElements = driver.findElements(By.className("\_4arz"));

for (j = 0; j < interestElements.size(); j++) {

try {

temp = interestElements.get(j).getText();

temp = temp.replaceAll("명", "").replaceAll(",", "");

if (temp.contains("천")) {

interestPerWord += Double.parseDouble(temp.replaceAll("천", "")) \* 1000;

} else if (temp.contains("만")) {

interestPerWord += Double.parseDouble(temp.replaceAll("만", "")) \* 10000;

} else {

interestPerWord += Double.parseDouble(temp);

}

} catch (Exception e) {

// TODO: handle exception

System.out.println(e.getMessage());

continue;

}

}

interestElements = driver.findElements(By.className("\_36\_q"));

for (j = 0; j < interestElements.size(); j++) {

try {

temp = interestElements.get(j).getText();

if (temp.contains("조회")) {

continue;

}

temp = temp.replaceAll("\\s", "").replaceAll("댓글", "").replaceAll("공유", "").replaceAll("개", "")

.replaceAll("회", "").replaceAll(",", "");

if (temp.contains("천")) {

interestPerWord += Double.parseDouble(temp.replaceAll("천", "")) \* 1000;

} else if (temp.contains("만")) {

interestPerWord += Double.parseDouble(temp.replaceAll("만", "")) \* 10000;

} else {

interestPerWord += Double.parseDouble(temp);

}

} catch (Exception e) {

// TODO: handle exception

System.out.println(e.getMessage());

continue;

}

}

words[i].addDegree(interestPerWord);

* 좋아요를 누른 사람 수가 담긴 태그와 댓글 수가 담간 태그, 공유 수가 담긴 태그에서 그 정보들을 가져와서 숫자를 추출해 객체에 저장한다.