

```

package com.sxt.regex.test;

import java.io.BufferedReader;
import java.io.InputStreamReader;
import java.net.URL;
import java.nio.charset.Charset;
import java.util.regex.Matcher;
import java.util.regex.Pattern;

/**
 * 网络爬虫链接
 * @author 江
 *
 */

public class WebSpiderTest {
    public static String getURLContent(String urlStr) {
        StringBuilder sb=new StringBuilder();
        try {
            URL url=new URL(urlStr);
            BufferedReader reader=new BufferedReader(new
InputStreamReader(url.openStream(),Charset.forName("gbk")));
            String temp="";
            while((temp=reader.readLine())!=null) {
                sb.append(temp);
            }
            return sb.toString();
        } catch (Exception e) {
            e.printStackTrace();
            return null;
        }
    }

    public static void main(String[] args) {

```

```
String destStr=getURLContent("http://www.163.com");
//
内容
Pattern p=Pattern.compile("<a[\\s\\S]+?</a>");    //取到的超链接的整个内
容
Pattern p=Pattern.compile("href=\"(.+?)\"");    //取到的超链接的整个内

Matcher m=p.matcher(destStr);
while(m.find()) {
    System.err.println(m.group());
    System.err.println(m.group(1));
}

}
}
```