

Piwik integration in WebLicht

Introduction

Piwik can be integrated in an web application in two manners:

1. As a Javascript snippet that registers page requests
2. Via a RESTful API

In WebLicht we were primarily interested in (2), since WebLicht is a single page (Zk-based) application. Moreover, we were interested in the web services that are used in a chain, such that we can have per-service statistics. For these reasons, we call Piwik in the WebLicht service chainer upon every service invocation.

Integration

Although calling the API directly is fairly simple, we decided to use a Java library that was available and recommended on the Piwik site¹. To report requests to Piwik, one simply creates a *SimplePiwikTracker* instance and sets the API URL and API key:

```
SimplePiwikTracker tracker = new SimplePiwikTracker(  
    PIWIK_ID_SITE, PIWIK_API_URL, request);  
tracker.setTokenAuth(PIWIK_TOKEN_AUTH);
```

The necessary information can be obtained as follows:

- The *request* object is a *HttpServletRequest* instance and is used to extract e.g. the clients IP address and user agent.
- The API key can be looked up by logging in to Piwik with the target account, clicking the triangle next to the user account and selecting *API*.
- The site ID is shown in the URL when looking up the API key.
- The URL for the test instance is: <https://stats.clarin.eu>

The *tracker* instance can then be used to report visits/calls/etc.:

```
URL trackURL = tracker.getDownloadTrackURL(  
    service.getUri().toString());  
tracker.sendRequest(trackURL);
```

This will register trackURL as a visited URL.

¹ <https://github.com/piwik/piwik-java-tracking>

Reverse proxy

When a reverse proxy is used (as is the case for WebLicht), the *HttpServletRequest* instance will contain the IP address of the reverse proxy server, rather than the client's IP address. Luckily, most reverse proxies (such as Apache) add the *X-FORWARDED-FOR* header, which contains the IP address of the original client.

Currently, the Java Piwik library does not use this header. So, some extra work should be done if the application is behind a proxy:

```
// Usually, requests to WebLicht are proxied. In such cases, we
// should be able to obtain the client IP through the
// 'X-Forwarded-For' header.
String forwardedFor = request.getHeader(X_FORWARDED_FOR);
if (forwardedFor != null) {
    // Apache mod_proxy documentation: Be careful when using these
    // headers on the origin server, since they will contain more
    // than one (comma-separated) value if the original
    // request already contained one of these headers.
    int sepIndex = forwardedFor.indexOf(',');
    if (sepIndex != -1) {
        forwardedFor = forwardedFor.substring(0, sepIndex);
    }
    LOGGER.debug("Forwarded for: {}", forwardedFor);
    newTracker.setIp(forwardedFor);
}
```

Normally, an API client **cannot** specify the IP address. To do this, an admin-level account should be requested from the Piwik administrators.

Result

After incorporating these changes, you can obtain request statistics such as those shown below in Piwik.

+ ws1-clarind.esc.rzg.mpg.de	12	13
+ weblicht.sfs.uni-tuebingen.de	372	18026
+ kaskade.dwds.de	12	12
+ dspin.dwds.de:8080	51	64
+ dllexdb.de	5	5
- clarin05.ims.uni-stuttgart.de	139	3832
/treetagger2008 	35	55
/treetagger	4	4
/RFTaggerMorph	26	3645
/rftagger	1	1
/cgi-bin/dspin/tokeniser4.perl	43	75
/cgi-bin/dspin/tei2tcf4.perl	1	1
/cgi-bin/dspin/tei2tcf3.perl	2	3
/cgi-bin/dspin/smor4.perl	1	1
/cgi-bin/dspin/bitpar4.perl	26	47
+ chopin.ipipan.waw.pl:8083	1	1