

Fakultät II: Department für Informatik
Projektgruppe 2011/2012

Installationshandbuch

Zur Anwendung *PlagTag*

Version 1.0



Betreuer:

Prof. Dr. Andreas Winter
M.Sc. Jan Jelschen

Mitglieder:

Tore Bierwirth
Christoph Gerken
Marion Gottschalk

Sieglinde Hahn
Maxim Klimenko

Björn Wolff
Christian Wübbeling

Oldenburg, den 30. September 2012

Inhaltsverzeichnis

1. Einleitung	1
1.1. Überblick und Ziel dieses Dokuments	1
2. Installation von PlagTag	2
2.1. Systemvoraussetzungen	2
2.2. Wichtige Empfehlungen	2
2.3. Installation	3
2.3.1. MySQL	3
2.3.2. Apache Tomcat	3
2.3.3. Deployment der Applikations-Archive und Start von Tomcat	4
2.3.4. PHP	4
2.3.5. PEAR	4
2.3.6. Pfade für Web-Frontend	5
2.4. Microsoft Bing-Konto	5
2.5. Testserver	5
2.5.1. Weiteres Vorgehen	5
A. Literaturverzeichnis	6
B. Glossar	7

1. Einleitung

Autor: Christian Wübbeling

Wir als Projektgruppe *Clone Busters* freuen uns, Ihnen diese Version der Software [PlagTag](#) als Ergebnis einer einjährigen Projektphase an der Universität Oldenburg präsentieren zu dürfen! Bei Fragen oder Anregungen wenden Sie sich bitte an die Mitglieder der Projektgruppe unter pg-clonebusters@uni-oldenburg.de.

1.1. Überblick und Ziel dieses Dokuments

Mithilfe der Software [PlagTag](#) haben Sie die Möglichkeit, textuelle [Plagiate](#) in Dateien aufzuspüren. Die Software ermöglicht es dabei, verschiedene Arten von [Plagiaten](#) in einer Auswahl an Dateiformaten erkennen zu können.

Dabei baut das Software-System auf einer [Komponenten](#)-Architektur auf, welche über Netzwerkschnittstellen kommunizieren. Dem Benutzer stellt sich die Software mittels eines einfach zu bedienenden [Webfrontends](#) dar, welches über gängige Web-Browser bedienbar ist.

Bitte beachten Sie unbedingt die Systemvoraussetzungen in Kapitel [2](#), um die Software für sich nutzen zu können. In den darauf folgenden Kapiteln erfahren Sie, wie Sie die Software-Fremdkomponenten auf Ihrem System einrichten und konfigurieren können. In der Folge fahren Sie bitte mit der Lektüre des Wartungshandbuches fort, um die Software selbst korrekt und ihren Bedürfnissen entsprechend zu konfigurieren.

Dieses Dokument richtet sich an Sie als Systemadministrator in der IT-Verwaltung Ihrer Organisation.

2. Installation von PlagTag

Autor: Christian Wübbeling, Christoph Gerken

In diesem Kapitel finden Sie wichtige Hinweise, um die Software installieren und konfigurieren zu können.

2.1. Systemvoraussetzungen

Bitte beachten Sie die folgenden Systemvoraussetzungen, um die Software einsetzen zu können:

- ein oder mehrere Computersysteme als Ausführungsplattform
- einen aktuellen Multicore-Prozessor
- mindestens 8GB Arbeitsspeicher
- ein Betriebssystem mit Unterstützung für Java 7 (1.7 JRE), siehe <http://www.oracle.com/>
- Java 7 (1.7) JRE, 64 Bit-Edition, siehe <http://www.oracle.com/>
- Apache Tomcat 7 - Applikationsserver, siehe <http://www.apache.org/tomcat/>
- Apache 2 - Webserver, siehe <http://www.apache.org/apache/>
- PHP 5.3 mit Pear-Erweiterungen (siehe Auflistung 2.3.5), siehe <http://www.php.org/>
- MySQL Community Server Version 5.5, siehe <http://www.mysql.com/>
- eine breitbandige Internetverbindung
- einen aktuellen Web-Browser, getestet auf Mozilla Firefox, Apple Safari und Google Chrome, siehe <http://www.mozilla.org/>, <http://www.apple.com/> und <http://www.google.com/chrome/>
- nur für Entwickler: Apache ANT, siehe <http://www.apache.org/ant/>
- nur für Entwickler: Eclipse Indigo SR2, siehe <http://www.eclipse.org/>
- nur für Entwickler: Apache Tuscany 1.6.2 Binaries + Source, siehe <http://www.apache.org/tuscany/>
- nur für Entwickler: Java 7 (1.7) JDK, 64 Bit-Edition, siehe <http://www.oracle.com/>
- optional: Einen clusterbasierten Rechnerverbund auf Unix-Basis

Gegebenenfalls reichen auch geringere Systemvoraussetzungen, eine Funktionalität kann hier jedoch nicht garantiert werden.

2.2. Wichtige Empfehlungen

Ihre Daten sind uns so wichtig wie Ihnen. Daher setzt die Software auf erprobte Industriestandards und Ihre Mitverantwortung. Durch Überspannungsschäden, Hardware-Defekte oder Diebstahl können Ihre Daten jederzeit unerwartet vernichtet werden oder verloren gehen. Sichern Sie die auf dem Server befindliche Datenbank mit Ihren wichtigen Kundendaten daher regelmäßig, indem Sie das dem mySQL-Server mitgelieferte Programm „mysqldump“ verwenden. Verlassen Sie sich nicht auf gelegentliche Sicherungen: Richten Sie automatische Sicherungen ein und heben Sie ältere Sicherungen einige Sicherungs-Perioden lang auf.

Installieren Sie regelmäßig Updates des mySQL-Dienstes, Java und Apache Tomcat, um bei Software üblicherweise auftretende Sicherheitsprobleme oder weitere Fehler zu beheben.

2.3. Installation

Die folgenden Installationsanweisungen werden für ein Windows-System beschrieben und wurden hierfür getestet. Sämtliche Software ist jedoch auf unter Linux oder Mac OS lauffähig, da die verwendeten Komponenten für alle verbreiteten Betriebssysteme zur Verfügung stehen. Hierbei kann die notwendigen Konfiguration jedoch abweichen.

Verwenden Sie zur Installation sowohl das der Dokumentation beiliegende Datenbank-Abbild als auch die Binärdateien im WAR-Format. Alle weiteren Softwareprogramme beziehen Sie bitte aus dem Internet.

2.3.1. MySQL

Installieren Sie die aktuelle Version des [MySQL Community Server](#), setzen Sie ein sicheres Passwort für den Benutzer `root` und erweitern Sie die Datei `my.ini` um die Zeile: `max_allowed_packet=64M`, um das Speichern von größeren Dateien zu ermöglichen. Spielen Sie nun das Datenbankimage `plagtag.sql` ein. Hierfür kann das `mysql-Kommandozeilen-Programm` verwendet werden (`mysql -u user -p passwort tabelle < plagtag.sql`) oder gängige Programme wie [PHPMyAdmin](#) oder [HeidiSQL](#).

Apache ANT, Tuscany und Eclipse

Hinweis: Die folgenden Komponenten sind nur für Entwickler erforderlich. Eclipse unterstützt Sie als Entwicklungsumgebung bei der Anwendungsentwicklung, Ant führt Kompilierungs- und sonstige Erstellungsaufgaben im Hinblick auf Anwendungsarchive aus, [Tuscany](#) ist als einzubindende Bibliothek erforderlich.

Installieren sie alle Softwarepakete in einem separaten Ordner. Achten Sie darauf, dass der Ordnerpfad keine Leerzeichen enthält. Danach setzen Sie folgende Systemvariablen anhand der von Ihnen gewählten Pfade:

- Systemvariable `JAVA_HOME` = z.B. `C:\Program Files\Java\jdk1.7.0_01` und
- Systemvariable `JRE_HOME` = z.B. `C:\Program Files\Java\jre7`
- Systemvariable `TUSCANY_HOME` = z.B. `C:\tuscany`
- Systemvariable `ANT_HOME` = z.B. `C:\ant`
- zur bestehenden Systemvariable `PATH` hinzufügen: z.B. `C:\ant\bin`
- Systemvariable `CATALINA_HOME` = z.B. `C:\tomcat`

Die Systemvariablen können Sie unter Windows 7 ändern: **Systemsteuerung - System - Erweiterte Einstellungen - Register Erweitert - Umgebungsvariablen**. Nach einer Änderung müssen Kommandozeilen-Fenster neu geöffnet werden.

2.3.2. Apache Tomcat

Apache Tomcat führt als Applikationsserver die eigentliche Software aus.

Installieren Sie die neueste Version von Apache Tomcat in einen Ordner ohne Leerzeichen im Pfad. Stellen Sie folgende Systemvariablen ein:

- Systemvariable `JAVA_HOME`, z.B. `C:\Program Files\Java\jdk1.7.0_07` und
- Systemvariable `JRE_HOME`, z.B. `C:\Program Files\Java\jre7`
- Systemvariable `CATALINA_HOME`, z.B. `C:\tomcat`

Die Systemvariablen können Sie unter Windows 7 ändern: **Systemsteuerung - System - Erweiterte Einstellungen - Register Erweitert - Umgebungsvariablen**. Nach einer Änderung müssen Kommandozeilen-Fenster neu geöffnet werden.

2.3.3. Deployment der Applikations-Archive und Start von Tomcat

Zum Start der Applikation sind die Applikations-Archive (WAR-Dateien) in den `webapps`-Unterverzeichnis des Tomcat-Ordners zu kopieren.

Einige Komponenten benötigen mehr Arbeitsspeicher als gewöhnlich, um korrekt arbeiten zu können. Unter Apache Tomcat muss daher die Datei `catalina.sh` (Unix/Linux) `setenv.bat` (Windows) um den Befehl `export CATALINA_OPTS=Xmx4g XX:MaxPermSize=512m` (Unix/Linux) bzw. `set CATALINA_OPTS=-Xmx4g -XX:MaxPermSize=512m` (Windows) ergänzt werden.

Danach kann Apache Tomcat über das Skript `startup.bat` im Unterverzeichnis `bin` gestartet werden. Wichtig: Stellen Sie vorher sicher, dass Sie eine 64Bit-Version von Java verwenden und die Umgebungsvariablen (siehe Auflistung oben) korrekt darauf angepasst wurden. Ist dies nicht der Fall, wird der Applikationsserver aufgrund der Speicherbeschränkungen von 32Bit-Systemen nicht starten.

2.3.4. PHP

Installieren Sie auf dem Apache Webserver ein aktuelles [PHP-Environment](#) (PHP 5.2 oder neuer). Diese muss in der Webserver Konfiguration des Apache-Webserver eingebunden werden. Je nach Betriebssystem- oder Webserver-Version passiert dies automatisch.

Damit das Webfrontend korrekt funktioniert, müssen in der Konfigurationsdatei von PHP (`php.ini`) einige Einstellungen korrigiert werden. Diese sind in der folgenden Liste mit den empfohlenen Werten aufgelistet.

- `max_execution_time` = 8000 (Ausführungszeit des Programms in Sekunden)
- `upload_max_filesize` = 20M (Maximale Größe eines Dateiuploads)
- `post_max_size` = 20M (Reguliert ebenfalls die maximale Größe einer Dateiübertragung)

2.3.5. PEAR

Pear ist eine Sammlung von Komponenten für die Programmiersprache [PHP](#). Diese werden standardisiert zu Verfügung gestellt und können über einen eigenen Installationsprozess per Kommandozeile installiert werden. Für die Nutzung des Webfrontend wurden folgende Bibliotheken verwendet.

- `SOAP-0.13.0` (WebService-Binding)
- `XML_Util-1.21` (XML-Verarbeitung)
- `Auth-1.6.4` (Benutzer-Authentisierung)
- `HTML_Table-1.8.3` (Seitenformatierung)
- `HTML_Quickform-3.2.13` (Seitenformatierung)
- `Log-1.12.7` (Logging)
- `HTTP_Session2-0.7.3-beta` (Session-Verwaltung)
- `Mail-1.20` (Mail-Versand)
- `Mail_Mime-1.8.5` (Mail-Erstellung)

Für die Installation muss auf der Kommandozeile für jede Komponente folgende Befehl eingegeben werden:

- `pear install [Komponente]`

Als Beispiel für die Komponente SOAP also „pear install SOAP“.

A. Literaturverzeichnis

- [LCF⁺11] LAWS, Simon ; COMBELLACK, Mark ; FENG, Raymond ; MAHBOD, Haleh ; NASH, Simon: *Tuscany SCA in Action*. Manning Publications Co., 2011
- [OAS] OASIS: *OASIS Service Component Architecture / Assembly*. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=sca-assembly. – Zugriff am 23.04.2012
- [PHP] PHP: *Migrating from PHP 4 to PHP 5.0.x*. <http://de.php.net/manual/de/migration5.php>. – Letzter Zugriff am 08.08.2012
- [Tec] TECHTERMS: *API*. <http://www.techterms.com/definition/api>. – Letzter Zugriff am 08.08.2012
- [w3s] W3SCHOOLS: *Introduction to SQL*. http://www.w3schools.com/sql/sql_intro.asp. – Letzter Zugriff am 08.08.2012

B. Glossar

API API steht für die Abkürzung *Application Programming Interface*. Dies ist ein Programmteil, welcher von einem Softwaresystem für ein anders **System** zur Verfügung gestellt wird [Tec].

AppID Die AppID ist eine alphanummerische Folge von Zeichen, welche bei der Anmeldung von **Bing** vergeben wird.

Bing Bing eine die kostenlose Suchmaschine von Microsoft, welche einen Zugriff auf ihre **API** über eine **AppID** ermöglicht.

Komponente Eine Komponente (hier: Software-Komponente) ist ein Software-Element, das konform zu einem Komponentenmodell (z.B. aus der UML) ist und über Schnittstellen mit anderen Komponenten verknüpft und ausgeführt werden kann. Komponentenbasierte Software wird z.B. mittels Web-Services, CORBA, Enterprise Java Beans (EJBs) oder COM entwickelt.

PHP Die Abkürzung PHP steht für *Hypertext Preprocessor* und ist eine Skriptsprache, welche hauptsächlich für die Erstellung von dynamischen Webseiten verwendet wird [PHP].

PlagTag PlagTag ist der Name der Plagiatserkennungssoftware.

Plagiat Ein Plagiat liegt dann vor, wenn jemand vorgibt, ein Werk selbst erstellt zu haben, obwohl die Inhalte ganz oder teilweise aus fremden Quellen stammen, die falsch oder unvollständig zitiert wurden. Je nachdem wie die Zitiertechnik falsch verwendet wurde, lassen sich die Plagiate zusätzlich genauer kategorisieren. Es existieren vielfältige Formen geistiger Leistung, z.B. Texte, Bilder oder Videos.

SCA Die Service Component Architecture (SCA) ist eine Sammlung an Spezifikationen, welche ein Modell einer Serviceorientierten Architektur (SOA) beschreiben. SCA basiert auf offenen Standards wie Web Services. SCA-Komponenten sind unabhängig von einer konkreten Technologie. [OAS].

SQL Die Abkürzung SQL steht für *Structured Query Language*. Es ist eine Sprache für Datenbankabfragen [w3s].

System Wird der System-Begriff ohne explizite Einschränkungen verwendet, so bezieht sich dieser immer auf das im Projekt zu entwickelnde Gesamtprodukt. Dieses teilt sich auf in eine zentrale Berechnungs-Komponente und ein angebundenes Webfrontend.

Tuscany Tuscany stellt ein kostenloses Framework zur Nutzung von **SCA** dar [LCF⁺11].

Webfrontend Das Webfrontend stellt ein System für den Anwender auf der Clientseite visuell dar. Es repräsentiert die Benutzeroberfläche (GUI).