**Martin-Luther-Universität Halle (Saale)**

****

Dokumentation der Studienleistung der Lehrveranstaltung Websuche- und Information Retrieval

Name: Erik Kinder

E-Mail: erik.kinder@student.uni-halle.de

Matrikelnummer:

Name: Jan Heinrich Reimer

E-Mail: jan.reimer@student.uni-halle.de

Matrikelnummer: 216204166

Name: Lukas Neef

E-Mail: lukas.neef@student.uni-halle.de

Matrikelnummer: 215246264

**Inhaltsverzeichnis**

[**1. Einleitung** 2](#_Toc534665191)

[**1.1. Problemstellung** 2](#_Toc534665192)

[**1.2. Zielstellung** 2](#_Toc534665193)

[**2. Motivation** 2](#_Toc534665194)

[**3. Vorgehensweise** 3](#_Toc534665195)

[**3.1. Crawler** 3](#_Toc534665196)

[**3.2. Parser** 3](#_Toc534665197)

[**4. Funktionsweise** 3](#_Toc534665198)

[**5. Schlussteil und Aussicht** 3](#_Toc534665199)

# **1. Einleitung**

## **1.1. Problemstellung**

Vorliegende Dokumentation beschreibt die Umsetzung der von der Lehrveranstaltung „Websuche- und Information Retrieval“ notwendigen Studienleistung. Für deren Erfüllung soll eine Applikation konzipiert werden, welche Suchanfragen verarbeitet und deren Treffer als Auflistung ausgibt. Die Anwendung der Suche soll in einem selber gewähltem Themengebiet erfolgen, welches seitens unserer Gruppe mit übergeordnetem Begriff „Libretto“ identifiziert wurde.

## **1.2. Zielstellung**

Das Ziel der vorliegenden Dokumentation soll es sein, eine Übersicht über die Umsetzung der Studienleistung zu geben und somit die weitere Bearbeitung des Resultats in der Zukunft zu ermöglichen.

# **2. Motivation**

Auf unserer Suche nach einem Gebiet für die gegliederte Suche von deren Inhalten, sind wir auf Libretto gestoßen. Diese erwiesen sich nach Recherche im Internet als eine Quelle von Informationen, welche noch kaum gegliedert vorzufinden sind. So stellten wir uns Szenarien wie den Besuch einer entsprechenden Aufführung vor, in der es nützlich wäre den Text der Oper, das Libretto, näher zu durchsuchen. So könnte es beispielsweise hilfreich sein, wenn der Text aus anderen Sprachen übersetzt wird, um ein besseres Verständnis für die Inhalte der Handlung zu bekommen. Auch würden Menschen mit einem schlechten Gehör die Möglichkeit bekommen, mit Hilfe des Textes der Handlung besser folgen zu können und nicht nur dem Geschehen auf der Bühne zuschauen zu müssen. Dies könnte auch anderweitig der Fall sein, wenn während der Aufführung ein bestimmter Teil nicht verstanden wurde, weil die Präsentation oder Aussprache sehr undeutlich war oder stark betont wurde. Mit Hilfe der Suche im Libretto könnten des Weiteren auch sämtliche Informationen wie Autor oder das Jahr der Erstaufführung recherchiert werden, was in der Vor- und Nachbereitung eines Besuches der Aufführung von Bedeutung ist. Die Möglichkeiten der Benutzung dieser Suchmaschine wären also vielfältig und könnten vielleicht sogar der Forschung der Librettologie zugutekommen.

# **3. Vorgehensweise**

## **3.1. Crawler**

Im ersten Schritt der Umsetzung ist eine umfassende Recherche nach Libretto und deren Texten im Internet erfolgt. Fündig wurden wir auf den Internetseiten operafolio.com und aria-database.com, welche beide über eine Vielzahl von Libretto verfügen und von uns daher für das weitere Vorgehen als hauptsächliche Quellen dienten. Um an die Inhalte beider Seiten zu gelangen und somit eine Grundlage für die spätere Suche zu schaffen, wurde im Anschluss ein Crawler konzipiert, welcher sämtliche Unterseiten erfasst und diese im HTML Format abspeichert. Dies wurde über eine Lösung in der PowerShell von Windows umgesetzt.

## **3.2. Parser**

asdgfgasdff

# **4. Funktionsweise**

# **5. Schlussteil und Aussicht**