

MASTER THESIS  
COMPUTER SCIENCE

# UIMA, Docker and Kafka

Buzzwords oder doch interessant?

*Simon Gehring*

Am Jesuitenhof 3  
53117 Bonn  
simon.gehring@fkie.fraunhofer.de  
Matriculation Number 2553262

At the  
RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN  
in cooperation with the  
FRAUNHOFER-INSTITUT FÜR KOMMUNIKATION,  
INFORMATIONSVERRARBEITUNG UND ERGONOMIE

supervised by  
Prof. Dr. Heiko RÖGLIN and Daniel TÖWS

February 16, 2018

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Motivation . . . . .	3
1.2	Basics . . . . .	3
1.2.1	UIMA . . . . .	3
1.2.2	Docker . . . . .	3
1.2.3	Hadoop . . . . .	3
1.2.4	Spark . . . . .	3
1.2.5	Kafka . . . . .	3
1.3	Problem . . . . .	3
1.3.1	Scaling UIMA . . . . .	3
1.3.2	Implementation Requirements . . . . .	3
1.4	Related Work . . . . .	3
1.4.1	Watson . . . . .	3
1.4.2	Something else that warrants another subsection . . . .	3
1.4.3	Maybe something not UIMA related? . . . . .	3
1.5	Outline . . . . .	3
<b>2</b>	<b>Implementation</b>	<b>4</b>
2.1	Concrete Application . . . . .	4
2.2	Documents -> Kafka . . . . .	4
2.3	Kafka -> Spark . . . . .	4
2.4	Spark -> UIMA . . . . .	4
2.5	UIMA -> Java . . . . .	4
2.6	Kafka -> Output . . . . .	4
2.7	Scaling all this . . . . .	4
<b>3</b>	<b>Evaluation</b>	<b>5</b>
3.1	Computation Speed . . . . .	5
3.2	Memory Usage . . . . .	5
3.3	Extensibility . . . . .	5
3.4	Maintainability . . . . .	5

<b>4</b>	<b>Summary</b>	<b>6</b>
4.1	The Judgement . . . . .	6
<b>5</b>	<b>Future Work</b>	<b>7</b>

# Chapter 1

## Introduction

### 1.1 Motivation

### 1.2 Basics

#### 1.2.1 UIMA

#### 1.2.2 Docker

#### 1.2.3 Hadoop

#### 1.2.4 Spark

#### 1.2.5 Kafka

### 1.3 Problem

#### 1.3.1 Scaling UIMA

UIMA-AS

UIMA-CPM

#### 1.3.2 Implementation Requirements

### 1.4 Related Work

#### 1.4.1 Watson

#### 1.4.2 Something else that warrants another subsection

#### 1.4.3 Maybe something not UIMA related?

### 1.5 Outline

## Chapter 2

# Implementation

2.1 Concrete Application

2.2 Documents -> Kafka

2.3 Kafka -> Spark

2.4 Spark -> UIMA

2.5 UIMA -> Java

2.6 Kafka -> Output

2.7 Scaling all this

## Chapter 3

# Evaluation

3.1 Computation Speed

3.2 Memory Usage

3.3 Extensibility

3.4 Maintainability

## Chapter 4

# Summary

### 4.1 The Judgement

## Chapter 5

# Future Work

UIMA (Unstructured Information Management Applications) ist toll! DKPro  
ist wundervoll! [SK99]



# Glossary

## **DKPro**

DKPro is a community of projects focussing on re-usable Natural Language Processing software <sup>7</sup>

## **Unstructured Information Management Applications**

UIMA are software systems that analyze large volumes of unstructured information in order to discover knowledge that is relevant to an end user. An example UIM application might ingest plain text and identify entities, such as persons, places, organizations; or relations, such as works-for or located-at. <sup>7</sup>

# Bibliography

- [SK99] Richard Matthew Stallman and Stefan Krempl. Software muß frei sein! Website, 1999. Online erhältlich unter <http://www.heise.de/tp/deutsch/inhalt/te/2860/1.html>; abgerufen am 8. Januar 2005.

## Eidesstattliche Erklärung

Hiermit versichere ich, Simon Gehring, dass ich die vorliegende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die Stellen meiner Arbeit, die dem Wortlaut oder dem Sinne nach anderen Werken und Quellen, einschließlich Quellen aus dem Internet, entnommen sind, habe ich in jedem Fall unter Angabe der Quelle deutlich als Entlehnung kenntlich gemacht. Dasselbe gilt sinngemäß für Tabellen, Karten und Abbildungen.

Unterschrift: \_\_\_\_\_  
Simon Gehring, Student  
Universität Bonn