

UIMA Tutorial and Developers' Guides

**Written and maintained by the Apache
UIMA™ Development Community**

Version 3.4.1

Copyright © 2006, 2023 The Apache Software Foundation

Copyright © 2004, 2006 International Business Machines Corporation

License and Disclaimer. The ASF licenses this documentation to you under the Apache License, Version 2.0 (the "License"); you may not use this documentation except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, this documentation and its contents are distributed under the License on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Trademarks. All terms mentioned in the text that are known to be trademarks or service marks have been appropriately capitalized. Use of such terms in this book should not be regarded as affecting the validity of the the trademark or service mark.

Publication date February, 2023

Table of Contents

1. Annotator & AE Developer's Guide	1
1.1. Getting Started	2
1.1.1. Defining Types	3
1.1.2. Generating Java Source Files for CAS Types	5
1.1.3. Developing Your Annotator Code	6
1.1.4. Creating the XML Descriptor	8
1.1.5. Testing Your Annotator	10
1.2. Configuration and Logging	13
1.2.1. Configuration Parameters	13
1.2.2. Logging	16
1.3. Building Aggregate Analysis Engines	20
1.3.1. Combining Annotators	20
1.3.2. AAEs can also contain CAS Consumers	23
1.3.3. Reading the Results of Previous Annotators	24
1.4. Other examples	26
1.5. Additional Topics	26
1.5.1. Annotator Methods	26
1.5.2. Reporting errors from Annotators	27
1.5.3. Throwing Exceptions from Annotators	28
1.5.4. Accessing External Resources	30
1.5.5. Result Specifications	37
1.5.6. Class path setup when using JCas	39
1.5.7. Using the Shell Scripts	39
1.6. Common Pitfalls	40
1.7. UIMA Objects in Eclipse Debugger	40
1.8. Analysis Engine XML Descriptor	41
1.8.1. Header and Annotator Class Identification	41
1.8.2. Simple Metadata Attributes	42
1.8.3. Type System Definition	42
1.8.4. Capabilities	42
1.8.5. Configuration Parameters (Optional)	43
2. CPE Developer's Guide	47
2.1. CPE Concepts	48
2.2. CPE Configurator and CAS viewer	49
2.2.1. Using the CPE Configurator	49
2.2.2. Running the CPE Configurator from Eclipse	53
2.3. Running a CPE from Your Own Java Application	54
2.3.1. Using Listeners	54
2.4. Developing Collection Processing Components	55
2.4.1. Developing Collection Readers	55
2.4.2. Developing CAS Initializers	60
2.4.3. Developing CAS Consumers	61
2.5. Deploying a CPE	63
2.5.1. Deploying Managed CAS Processors	65
2.5.2. Deploying Non-managed CAS Processors	66
2.5.3. Deploying Integrated CAS Processors	67
2.6. Collection Processing Examples	68
3. Application Developer's Guide	71
3.1. The UIMAFramework Class	71
3.2. Using Analysis Engines	71
3.2.1. Instantiating an Analysis Engine	72

3.2.2. Analyzing Text Documents	72
3.2.3. Analyzing Non-Text Artifacts	73
3.2.4. Accessing Analysis Results	73
3.2.5. Multi-threaded Applications	74
3.2.6. Multiple AEs & Creating Shared CASes	76
3.2.7. Saving CASes to file systems or general Streams	77
3.3. Using Collection Processing Engines	80
3.3.1. Running a CPE from a Descriptor	80
3.3.2. Configuring a CPE Descriptor Programmatically	80
3.4. Setting Configuration Parameters	82
3.5. Integrating Text Analysis and Search	83
3.5.1. Building an Index	83
3.6. Working with Remote Services	86
3.6.1. Deploying as a Vinci Service	86
3.6.2. Calling a UIMA Service	88
3.6.3. Restrictions on remotely deployed services	88
3.6.4. The Vinci Naming Services (VNS)	89
3.6.5. Configuring Timeout Settings	91
3.7. Increasing performance using parallelism	93
3.8. Monitoring AE Performance using JMX	94
3.9. Performance Tuning Options	95
4. Flow Controller Developer's Guide	97
4.1. Developing the Flow Controller Code	97
4.1.1. Flow Controller Interface Overview	97
4.1.2. Example Code	98
4.2. Creating the Flow Controller Descriptor	100
4.3. Adding Flow Controller to an Aggregate	101
4.4. Adding Flow Controller to CPE	102
4.5. Using Flow Controllers with CAS Multipliers	102
4.6. Continuing the Flow When Exceptions Occur	103
5. Annotations, Artifacts & Sofas	105
5.1. Terminology	105
5.1.1. Artifact	105
5.1.2. Subject of Analysis — Sofa	105
5.2. Formats of Sofa Data	105
5.3. Setting and Accessing Sofa Data	106
5.3.1. Setting Sofa Data	106
5.3.2. Accessing Sofa Data	106
5.3.3. Accessing Sofa Data using a Java Stream	106
5.4. The Sofa Feature Structure	107
5.5. Annotations	107
5.5.1. Built-in Annotation types	107
5.5.2. Annotations have an associated Sofa	108
5.6. AnnotationBase	108
6. Multiple CAS Views	109
6.1. CAS Views and Sofas	109
6.1.1. Naming CAS Views and Sofas	109
6.1.2. Multi/Single View parts in Applications	110
6.2. Multi-View Components	110
6.2.1. Deciding: Multi-View	110
6.2.2. Multi-View: additional capabilities	110
6.2.3. Component XML metadata	110
6.3. Sofa Capabilities & APIs for Apps	111

6.4. Sofa Name Mapping	111
6.4.1. Name Mapping in an Aggregate Descriptor	112
6.4.2. Name Mapping in a CPE Descriptor	112
6.4.3. CAS View received by Process	113
6.4.4. Name Mapping in a UIMA Application	113
6.4.5. Name Mapping for Remote Services	114
6.5. JCas extensions for Multiple Views	114
6.6. Sample Multi-View Application	114
6.6.1. Annotator Descriptor	115
6.6.2. Application Setup	115
6.6.3. Annotator Processing	115
6.6.4. Accessing the results of analysis	116
6.7. Views API Summary	117
7. CAS Multiplier	119
7.1. Developing the CAS Multiplier Code	119
7.1.1. CAS Multiplier Interface Overview	119
7.1.2. Getting an empty CAS Instance	120
7.1.3. Example Code	120
7.2. CAS Multiplier Descriptor	123
7.3. Using CAS Multipliers in Aggregates	124
7.3.1. Aggregate: Adding the CAS Multiplier	124
7.3.2. CAS Multipliers and Flow Control	125
7.3.3. Aggregate CAS Multipliers	126
7.4. CAS Multipliers in CPE's	127
7.5. Applications: Calling CAS Multipliers	127
7.5.1. Output CASes	127
7.5.2. CAS Multipliers with other AEs	128
7.6. Merging with CAS Multipliers	129
7.6.1. CAS Merging Overview	129
7.6.2. Example CAS Merger	129
7.6.3. SimpleTextMerger in an Aggregate	131
8. XMI & EMF	133
8.1. Overview	133
8.2. Converting an Ecore Model to or from a UIMA Type System	133
8.3. Using XMI CAS Serialization	134
8.3.1. Character Encoding Issues with XML Serialization	134
9. Managing different TypeSystems	137
9.1. Annotators, Type Merging, and Remotes	137
9.2. Supporting Remote Annotators	137
9.3. Type filtering support in Binary Compressed Serialization/Deserialization	137
9.4. Remote Services support with Compressed Binary Serialization	138
9.5. Compressed Binary serialization to/from files	138

[Click here to download full PDF material](#)