**Natural Language Processing for PDF/TIFF/Image Documents  
SEGMENTATION MODULE  
High Precision Natural Language Processing  
Technical Specification, v1.3**

# 1 Segment

## 1.1 Segment Overview

The segment NLP preprocessor contains the following primary classes, and their relationships:

* X

**Columns**

**Segment (Base Class)**

**Paragraphs**

**Headings**

**List**

**Fig. 1a High Level view of Segment Class Object Relationships**

## 1.2 Segment Properties

The segment class contains the following properties:

* TBD

**TBD**

**Segment (Base Class)**

**Fig. 1b Properties of the Segment Class**

**stopwords**

**sentiment**

## 1.3 Segment Overridden Operators

The following operators have their implementations (inherited from the base Object) class overridden:

* TBD

**Overridden Operators**

**TBD**

**Segment (Base Class)**

**TBD**

**Fig. 1c Overridden Operators of the Segment Class**

## 1.4 Segment Private Methods

The segment class contains of following private methods:

* TBD

**Fig. 1d Private Methods of the Segment Class**

## 1.5 Segment Public Methods

The words class contains of following public methods:

* There are no public methods.

## APPENDIX I: Updates

*v1.4*

# APPENDIX II: Anticipated Engineering

The following has been identified as enhancement/issues to be addressed in subsequent update:

1. Add support for splitting dual column pages.
2. Add support for splitting paragraphs.
3. Add support for splitting headings.

**Proprietary and Confidential Information  
Copyright ©2018, Epipog, All Rights Reserved**