

Bates Numbering in PDF Documents

Introduction

CONTENTS

[Introduction](#) 1

[Bates Numbering in PDF](#) 1

[Locating Bates Numbers in PDF Documents](#) 3

Bates Numbering is a method of indexing legal documents for easy identification and retrieval. Acrobat 8.0 allows users to apply Bates numbers to one document or to a whole set of documents, and they may apply more than one Bates Numbering sequence to a single or multiple documents. Document sets may also be given a unique set of numbers or prefixes which might include text such as a set number, case number, or firm name.

Third-party software developers need to be able to extract the Bates Numbering information for document management purposes as well as to display the numbers to users. This document explains the methods for locating the Bates numbers and describes the syntax so the numbers can be extracted from PDF documents.

Bates Numbering in PDF

Previous versions of Acrobat allowed users to add headers and footers to a document to accommodate Bates Numbering, but there was no support for leading zeros, and it was not possible to number sequentially across multiple documents. But Acrobat 8 addresses those issues and provides a robust UI for a better user experience.

Basic Bates Number Syntax

Acrobat adds Bates numbers as a page header or footer, represented as XML data in a PieceInfo dictionary, which is generally used to hold private application data. In this case, it is used by Acrobat to store a page object or form XObject dictionary.

Bates numbers may include a prefix and a suffix, using any alphanumeric characters chosen by the user. The following is an example of the XML data for a document that uses six-digit numbers, with a prefix of "ADBE" and a suffix of "DRAFT":

```
<Bates start="1" ndigits="6" prefix="ADBE" suffix="DRAFT"/>
```

The above statement is part of XML data that specifies the Bates numbers as part of a document header or footer. The footer or header

data specifies all appearance attributes such as position, font, point size, and color. The XML data is contained in a PiecelInfo dictionary that is stored in the data for each page in the document. The resulting page header for the first page would look like:

```
ADBE000001DRAFT
```

If the user chooses not to use a prefix or suffix, the resulting XML data omits those parts, and the XML data would appear as follows:

```
<Bates start="1" ndigits="6"/>
```

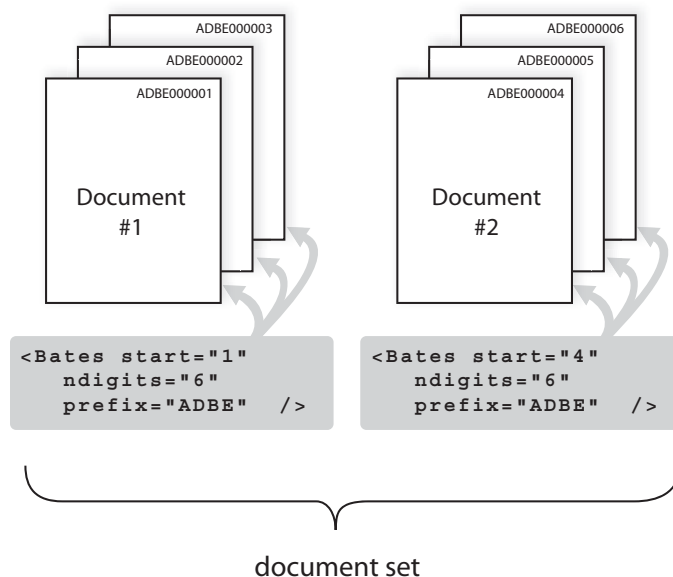
The Bates page numbers for each page are not explicitly stored in the PDF document. The XML data specifies the base number at which the numbering begins, and the number of digits in each number. For the above example, the data would look like:

```
<Bates start="1" ndigits="6" prefix="ADBE" suffix="DRAFT"/>
```

When used for a single document, the above statement would be stored for each page, and each instance of it would be identical. That is, only the start page number for the entire document is stored for each page. The viewing or managing application is responsible for calculating and displaying the correct number for all other pages.

Numbering for Document Sets

When Bates numbers are applied to a set of documents, the XML data is similar for each document, but each document's start value will be different, as shown in the following diagram for two documents with three pages each:



The diagram shows that for each document, identical XML data is stored for each page of that document, but the Bates Numbering parameters, such as start value, could carry different values from document to document.

Because multiple Bates Numbering sequences may be applied to any document, software attempting to extract Bates numbers should make sure that they locate all instances.

Locating Bates Numbers in PDF Documents

There are two methods for locating Bates numbers in a PDF document:

- An Acrobat plug-in can be developed to use the Acrobat API to locate and extract the Bates number information. The Acrobat SDK contains a sample plug-in: ExtractBatesInfo, which demonstrates how to use the API to work with Bates numbers in PDF. The Acrobat SDK is available at:
<http://www.adobe.com/devnet/acrobat/>
- PDF files can be scanned for the XML tag <Bates....>, and then an XML parser can be used to extract the needed information. The XML data is not compressed like other objects in a PDF document, so the data can be found without having to decompress data streams and parse the PDF.

CAUTIONS:

- Using Acrobat's Examine Document feature causes the PieceInfo dictionary to be removed, after which the Bates Numbering information cannot be extracted.
- When pages are extracted from a document having Bates Numbering, the numbers are correct. However, Bates Numbers computed using Bates Numbering parameters from the extracted pages will not match the original Bates Numbers displayed on the pages. To make them consistent, it is necessary to know the page index of the extracted pages in the original document. Currently, the ability to do that is not supported in Acrobat 8.0.

Copyright 2006–2007 Adobe Systems, Incorporated. All rights reserved.

Adobe Systems Incorporated

345 Park Avenue, San Jose, CA 95110-2704 USA

<http://www.adobe.com>

Adobe, the Adobe logo, Acrobat, Adobe LiveCycle, and Reader are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. Mac OS is a trademark of Apple Computer, Inc., registered in the United States and other countries. Linux is a registered trademark of Linus Torvalds. Microsoft, Windows, and Word are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. Unix is a registered trademark of The Open Group. All other trademarks are the property of their respective owners.

15 January 2007