



Maintenance and Reengineering of Information Systems

Presentation of the Project



Team



Team name: **Phaethon**

Team Members: **1**

Name: **Panagiotis**

Surname: **Toumasis**

ID (AM): **8000131**

email: **ptoumasis@mail.gr**

Project Details (1)



Project name: **PDFBox**

Project web site URL:

<http://sourceforge.net/projects/pdfbox>,
www.pdfbox.org

Topic: **Dynamic Content, Indexing/Search, Text Processing**

Development Status: **4 – Beta**

Implementation – Programming language: **Java**

21/4/2004 Platform: **OS Independent**

Maintenance and Reengineering
of Information Systems

Project Details (2)



Intended Audience: **Developers**

License: **BSD Licence (similar to MIT Licence)**

Development team size: **2 (Ben Litchfield, Bob Dickinson)**

Approximate source code size:

**30 packages, 139 classes,
more than 10000 lines of java code**

Registration Date: **7/4/2003**

Project Details (3)



Short description of the project:

- Ⓢ PDFBox is a Java PDF Library.
- Ⓢ This project will allow access to all of the components in a PDF document.
- Ⓢ More PDF manipulation features will be added as the project matures.
- Ⓢ It ships with a utility to take a PDF document and output a text file.

Project Versions



@PDFBox – 0.6.1

Released 11/4/2003, 7:19, Ben Litchfield

@PDFBox – 0.6.2

Released 18/4/2003, 21:40, Bob Dickinson

@PDFBox – 0.6.3

Released 14/9/2003, 6:57, Ben Litchfield

@PDFBox – 0.6.5

Released 8/3/2004, 19:28, Ben Litchfield

Some Bugs



Request ID	Summary	Date	Assigned to	Submitted by
927793	org.pdfbox.encrypted.DecryptDocument	1/4/2004,	nobody	Pcwizard_01
926998	Class Cast Exception During appendDocument	12:15 31/3/2004,	benlitchfield	valnifj
921378	IOException: expected='endobj'	13:07 22/3/2004,	nobody	nobody
921373	No 'ToUnicode' and no 'Encoding' for Font	13:18 22/3/2004,	nobody	nobody
920900	PDFTextStripper hangs on extraction	13:11 22/3/2004, 2:48	nobody	Hal_9000
919215	IOException in LucenePDFDocument.getDocument	18/3/2004, 17:40	nobody	nobody

My contribution (non-binding)



Short description of my planned contribution (non-binding):

- Ⓢ Debugging (some of the 21 bugs, if possible!).
- Ⓢ Including some viewing capabilities in the PDF Viewer (a GUI application) that show the internal structure of a PDF document.

[This is a task that the members of the team want to add to one of the next versions of the PDFbox.]