

FIRST - Bug #3453

Exception extracting text from doc11.pdf

07/16/2008 07:05 AM - Ryan McFall

Status:	Resolved	Start date:	07/16/2008
Priority:	Normal	Due date:	
Assignee:	Ryan McFall	% Done:	0%
Category:	parser	Estimated time:	0.00 hour
Target version:	Unspecified	Spent time:	0.00 hour
Bugzilla-Id:	3453		

Description

When parsing doc11.pdf from the private/samples directory, calling PDFExtract.extract gives the following stack trace:

```
java.util.NoSuchElementException
at java.util.LinkedList$ListItr.next(Unknown Source)
at edu.msu.first.parser.extract.PDFExtract.stripHeaders(PDFExtract.java:931)
at edu.msu.first.parser.extract.PDFExtract.extractContent(PDFExtract.java:206)
at edu.msu.first.parser.extract.PDFExtract.extract(PDFExtract.java:136)
```

History

#1 - 07/16/2008 10:22 AM - Bridger Hamilton

PDFExtract.stripHeaders now checks to make sure there is another line of text before checking to see if it's a header.

#2 - 03/27/2013 02:23 PM - Redmine Admin

Original Bugzilla ID was 3453