

## FIRST - Bug #3453

### Exception extracting text from doc11.pdf

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

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