

Document Parsing

Given that vendors follow up with Order Confirmations (OC) for every purchase order that is being submitted to them, parse and extract information from the attached document. OC is a pdf document, the result of the parsing should be a hash with the following key-value pairs:

- confirmed_on: <DateTime> [10/08/14 12:39]
- vendor_name: <String> "Maintenance-USA"
- order_number: <String> "12345678"
- status: <String> "pending"
- po_number: <String> "22551-1"
- total: <Float> 713.79
- line_items: <Array>
 - line_number: <Fixnum> 1
 - item_number: <String> "105350"
 - description: <String> "AMANA ELEC DRYER WHITE"
 - quantity: <Fixnum> 1
 - unit_price: <Float> 374.79
 - line_number: <Fixnum> 2
 - item_number: <String> "105342"
 - description: <String> "AMANA TOP LOAD WASHER"
 - quantity: <Fixnum> 1
 - unit_price: <Float> 339.00

Notes: Please include a README on how to run the script. The output should be a hash where the value types are indicated in angle brackets (<>). Tests are a plus. Also added bonus is to use ActiveRecord models to store relational data in. The result of parsing in that case would still be a hash but the data would come from objects instead.