**Parsing CV**

* pdfminer – python library for reading pdf files. (<https://pypi.python.org/pypi/pdfminer/>)
* another library that might be useful: <https://pypi.python.org/pypi/docx>
* will need to use regex
* Research suggests we need to use Natural Language Processing/ Text Mining.
* 3 models of CV parsing: <http://www.daxtra.com/2014/02/26/types-of-parser-and-how-they-work/>
* I think the biggest task is going to reading the data from the lines (various file formats!)
* The key things we want to extract – do we want to add or remove any, can we rank them so I can priorities:
  + Full Name
  + Address
  + School/University
  + Subjects/Grades/Degree
  + Job titles/ Company/ Duration
  + Languages
  + Hobby
* I found some GitHub pages:
  + <https://github.com/ssharad/resume-parser-in-python>
  + <https://github.com/lakshmanaram/Resume-parser>
  + <https://github.com/antonydeepak/ResumeParser>
  + I still need to make sense of them
* CV parsing software available: <http://socialcompare.com/en/comparison/resume-cv-parsers>