**Statistics**

**Total PDF’s**: 4754  
**OCR Saves (# of OCRS worked):** 532/1195  
**Total unable to extract:** 663 **Total Keywords:** 11463  
**Merging Results**: 58 keywords altered  
  
  
Notes:

* Keywords not cleaned (besides some manual cleaning)
* Vocabularies downloaded 159 json / 163 xml/rdf from Research Vocab Australia. However many overlaps between json and rdf files. Also many databases not viable eg examples (still used).
* Merging does not yet have language checks
* PDF gathering only multiprocessed, not multithreaded
* Merging only does simple judging based on number of altLabels gained.
* Found a good Wikipedia dump, that could be parsed to harvest broader links, <https://wiki.dbpedia.org/develop/datasets>.
* Was difficult to find any useful SKOS vocabularies to download