QA with Wiki: improving information retrieval \textit{and} machine comprehension

Rohan Sampath, Puyang Ma

**Motivation:**
EM pipeline is much lower than $EM_{retriever} \times EM_{reader}$ because:
(a) Flawed $EM_{retriever}$ score, (b) DocReader trained on SQuAD, tested on Wikipedia articles, and (c) True lack of fit between Retriever and Reader (since optimization is not done across entire pipeline)

**Conclusions**
- PageRank – limited improvement on Doc Retriever and overall pipeline performance
- BERT – significant improvement in Reader performance; some evidence to suggest that improvement gains not as high when answers are long
- Immediate next step – how does BERT affect $\gamma$ and how much does it improve overall pipeline performance by?

**References**