

E1246 Natural Language Understanding

Pathapati Abhishek

SR.NO.14594

pathapatia@iisc.ac.in

1.

Corpus	2gram(perplexity)	3gram(perplexity)	4gram(perplexity)
Brown	175.21	58.33	36.82
Gutenberg	201.22	79.14	62.1
Brown and Gutenberg(Test case Gutenberg)	337.0	117.25	89.46
Brown and Gutenberg(Test case brown)	667.51	345.23	126.4

The above tells us the perplexity for the given corpora. It shows that the values for the Brown corpus 4gram is better.

2. The file generates using 4grams generated using Brown corpus. Brown corpus is used because the English used is pertinent to the current usage. Some of the sentences generated using this file are:
- (a) superiority over those who invited them to the roof and energy is based upon formulas that
 - (b) if given in large minorities of germans are simply allowing life magazine to the roof and
 - (c) the weeks went ahead of him to offer the wechsler intelligence scale of compulsivity was