

Stemed=False			
Attribute Names	n_features (A)	sparcity(%)	
Category			
ALLLang	16	2.796875	
(Pure Eng. w) OnlyEng	41	2.4390243902	
Description			
AllLang	79510	0.1030793611	
(Pure Eng. w) OnlyEng	79510	0.1030793611	
Name			
AllLang	9204	0.0305747501	
(Pure Eng. w) OnlyEng	9204	0.0305747501	
Stemmed=True	All	OnlyEng	
Category	2.4390243902	2.796875	
Description	0.4011256824	1.1027593819	
Name	0.0337264667	0.0581003236	
Stemmed=False	All	OnlyEng	
Category	2.4390243902	2.796875	
Description	0.3611290556	0.7403105073	
Name	0.0308842479	0.0475760783	
PLOT - 2			
	Category:All	Category:OnlyEn	Description:All
Unstemmed	2.4390243902	2.796875	0.3611290556
Stemmed	2.4390243902	2.796875	0.4011256824

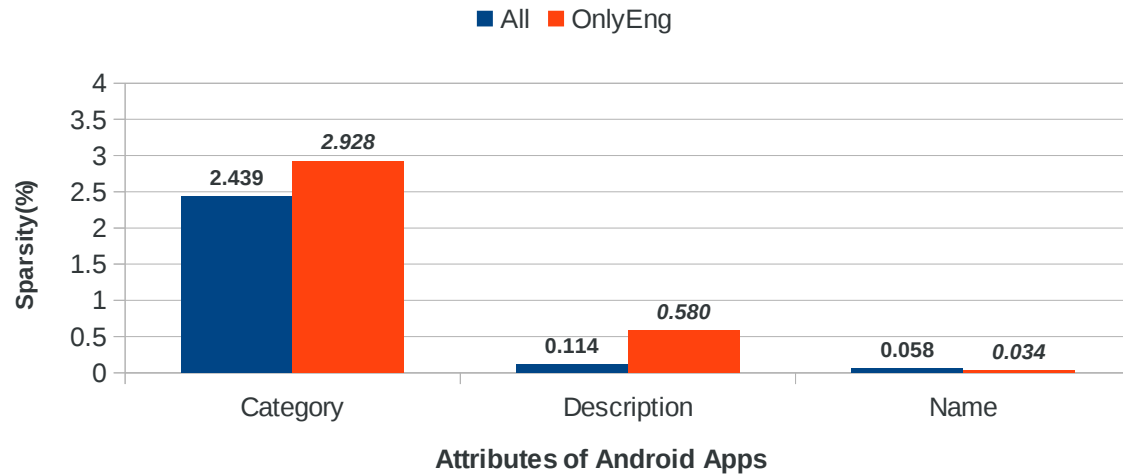
Stemmed Vectorization		n_samples	Remark	Stop Words
n_features	sparcity(%)			
		10K		Chinese
				English
41	2.439024	10K		Italian
16	2.796875	10K		Espanoil
				Arabic
67006	0.1136739993	5K		Swedish
9380	0.5800746269	5K		Japanese
				Finish
	0.0581003236	10K		Latin
3090	0.0337264667	10K		French
				German
				Total: 35

DescripticName : ALL Name : OnlyEng

0.7403	0.0308842479	0.0475760783
1.1028	0.0337264667	0.0581003236

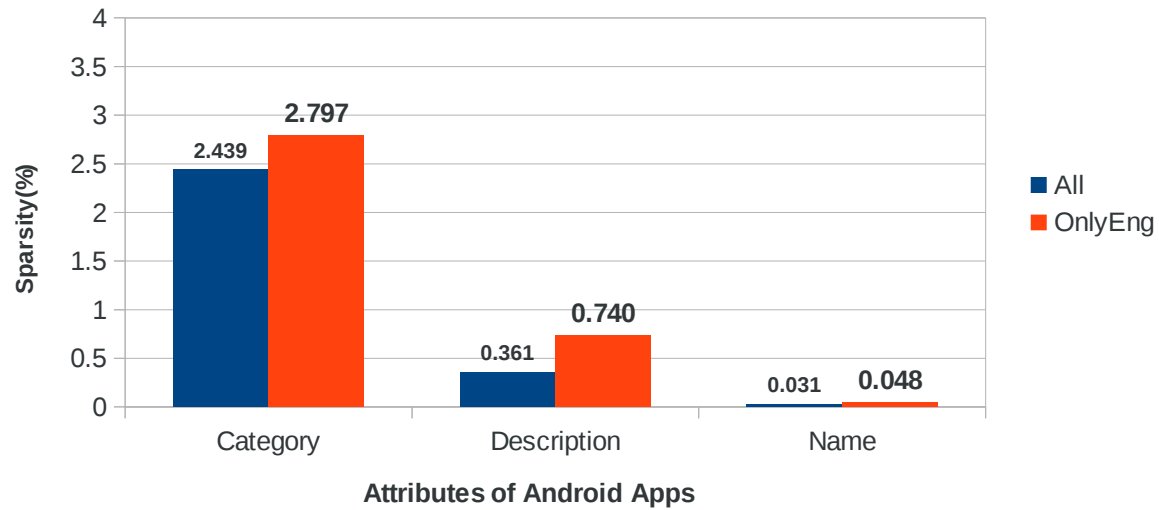
Sparsity of Attrbutes By Language Sets

Pure English vs All Languages Terms (Stemed)



Sparsity of Attributes By Language Sets

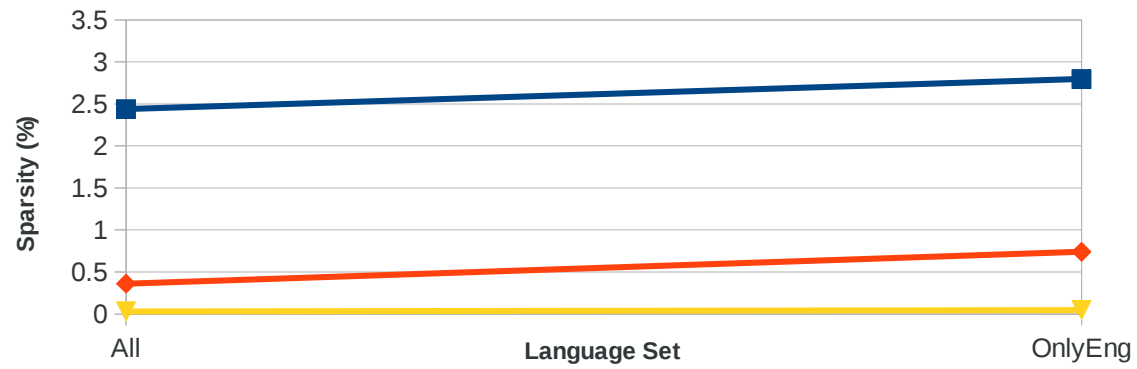
Pure English vs All Languages Terms (Unstemmed)



Sparsity Across Language Sets

UnStemmed

■ Category ■ Description ■ Name



Sparsity: Stemming Vs Unstemming Vectorization

