Project Notes

Tasks:

1. Improve Framework   che
2. Inference   Sanrav
3. Hashtables
4. Learning   Benjamin

5. Preprocessing : • Stemming for T
                            • Stop words
                            • Tokenize
6. Choose dataset, download it

ck models for sanity   Fred

(Represent docs as list of integers)

opic Modelling doesn't work.
            Cristina

\ | |

One week for those

Corpus:

2 steps : 1) Transfer text repo
2) Load that from

Next steps: • Measuring framew
• Profile the code

tasks !

. to some vector rep.

a file

ork

for hot spots

Could be done
e.g Python, b
much ha

e in

t not

neter in C