

CS15-319 / 15-619

Cloud Computing

Recitation 15

April 30th, 2013

Announcements

- Checkpoint Quiz Unit 5, due on:
 - Friday May 3rd at midnight
- Project 4, Part c, due on:
 - Friday May 3rd at midnight

Announcements

- **Open up S3 location of hand ins:**
 - Give access to your S3 bucket to:
 - public
 - onlinecloudcomputingcourse@gmail.com
 - You could lose credit or be penalized otherwise
 - See Piazza Post on how to open up your handin directory
- Encounter a general bug:
 - Post on Piazza
- Encounter a grading bug:
 - Post Privately on Piazza
- Post feedback on OLI



Announcements

- **Amazon Account Closure**
 - After project deadline
 - Amazon accounts will be unlinked from the course payment account
 - Shut down instances and services or your card on file may get charged.
- **Course Survey**
 - Anonymous, Web form link will be dispatched by email.
 - Students who complete will receive **2% bonus** points boost on final grade.

HBase Tips

- Updated hbase.properties file
 - <https://s3.amazonaws.com/15-319-s13/proj4/proj4-hbase.properties>
 - Use this properties file if you are having trouble with the previous version.

New Modules

- Unit 5 – Distributed Programming and Analytics Engines for the Cloud
 - Introduction to Distributed Programming for the Cloud
 - Distributed Analytics Engines for the Cloud: MapReduce
 - Distributed Analytics Engines for the Cloud: Pregel
 - Distributed Analytics Engines for the Cloud: **GraphLab** 
 - GraphLab
 - Data Structure and Graph Flow
 - The Architectural Model
 - The Programming Model
 - The Computation Model
 - Fault Tolerance
- Distributed Programming and Analytics Engines for the Cloud : **Summary**
 - Checkpoint Quiz 

Project 4, Part c

- Project 4, Part a
 - MapReduce
 - Project 4 Survey
- Project 4, Part b
 - Input Text Predictor: NGram Generation
- Project 4, Part c
 - Input Text Predictor: Language Model and User Interface

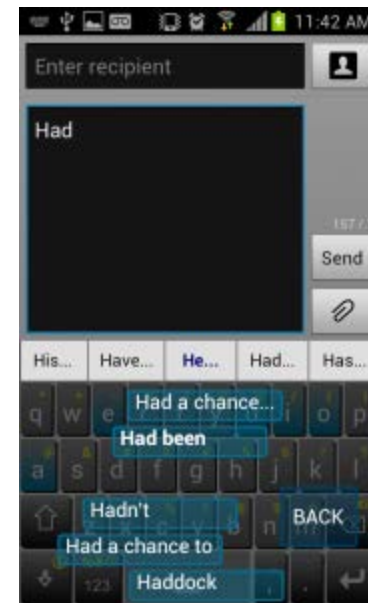


Recap Input Text Prediction

- Construct an Input Text Predictor

wiki		Advanced Search
wikipedia	250,000,000 results	Preferences
wikipedia encyclopedia	16,300,000 results	Language Tools
wiki answers	24,400,000 results	
wikimapia	12,000,000 results	
wikihow	1,780,000 results	Slovenija
wikiquote	3,280,000 results	
wikispaces	7,800,000 results	
wikitavel	2,270,000 results	
wikimedia	55,700,000 results	
wikipedia dictionary	20,300,000 results	
	close	

Google Suggest



WordLogic iKnowU keyboard

How to Construct an Input Text Predictor?


~~1. Given a language corpus~~

- Project Gutenberg (2.5GB, already on S3)
- English Language Wikipedia Articles (30GB, on S3 soon)

~~2. Construct an n-gram model of the corpus~~

- An n-gram is a phrase with n words.
- For example a set of 1,2,3,4,5-grams with counts:
 - this 1000
 - this is 500
 - this is a 125
 - this is a blue 60
 - this is a blue house 20

How to Construct an Input Text Predictor?

3. Build a statistical language model that contains the probability of a word appearing after a phrase 

$$- \Pr(is|this) = \frac{Count(this\ is)}{Count(this)} = \frac{500}{1000} = 0.5$$

$$- \Pr(a|this\ is) = \frac{Count(this\ is\ a)}{Count(this\ is)} = \frac{125}{500} = 0.25$$

4. Store and index the words and their probabilities to use in an application 

Discussions

- Your questions...

Upcoming Deadlines

- Unit 5:

Unit 5: Distributed Programming and Analytics Engines for the Cloud

Module 21: Distributed Analytics Engines for the Cloud: GraphLab

Module 22: Distributed Programming and Analytics Engines for the Cloud: Summary

Quiz 5: Distributed Programming and Analytics Engines for the Cloud

[Checkpoint](#)

Available Now

Due 5/3/13 11:59 PM



- Project 4

Project 4

Module 34: Input Text Predictor : Language Model and User Interface

Language Model Generation

[Checkpoint](#)

Available Now

Due 5/3/13 11:59 PM

