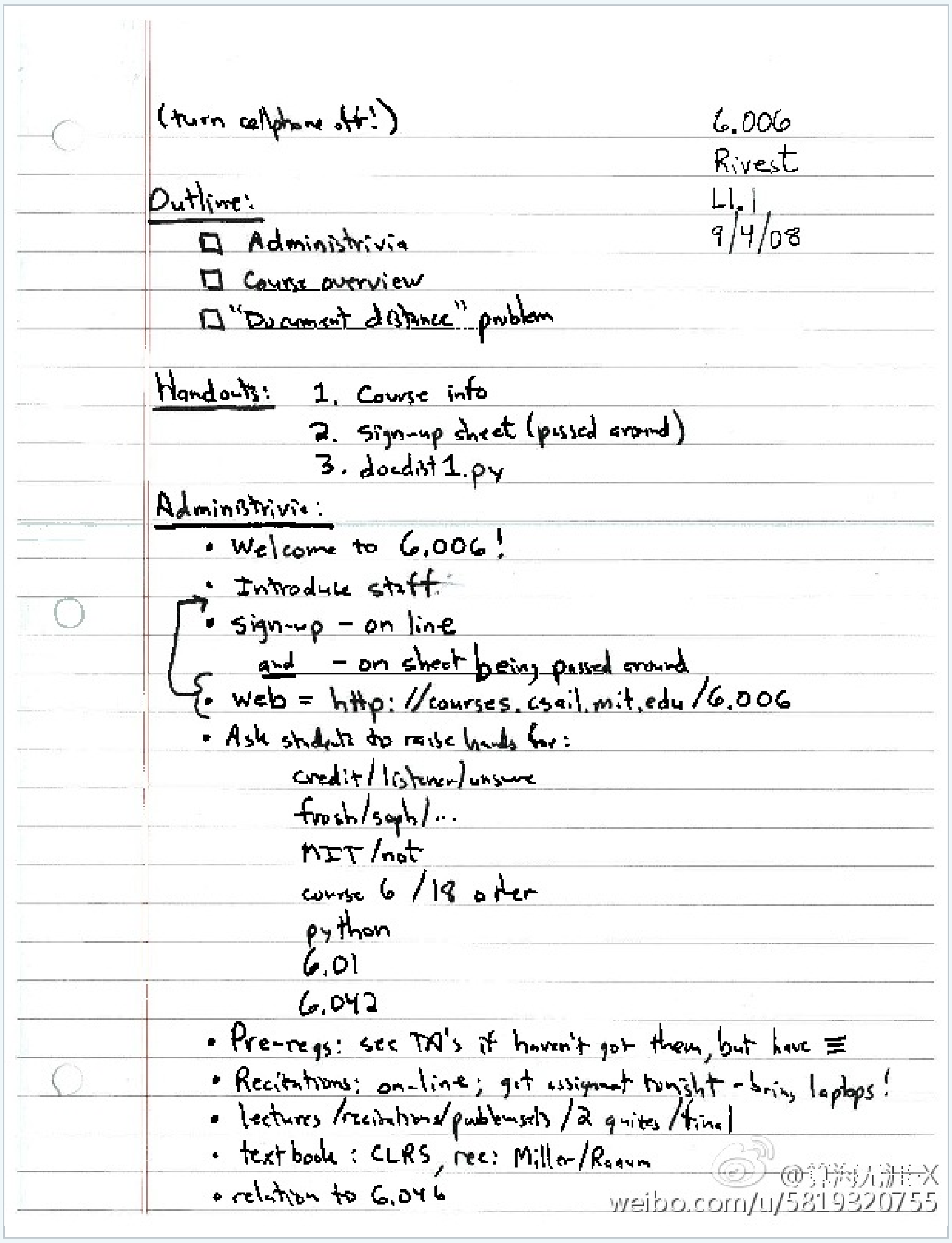
ILDER

@保存到为知笔记 // @网路冷眼 :转发微博

@算海无涯-X

#CS名校算法资源# MIT 6.006算法课考古贴: http://t.cn/RqdvHYu 收录了从2007年至今的课程资源，有两个发现：一是他们一直使用手写讲义并且传承悠久；二是从2007年开始就已经使用Python作为描述语言。附图是RSA密码算法字母R的那位大神Rivest的2008年珍贵手写讲义摘选： http://t.cn/RqdvHYn



(turn cellphone off!)

6.006
Rivest
L1.1
9/4/08

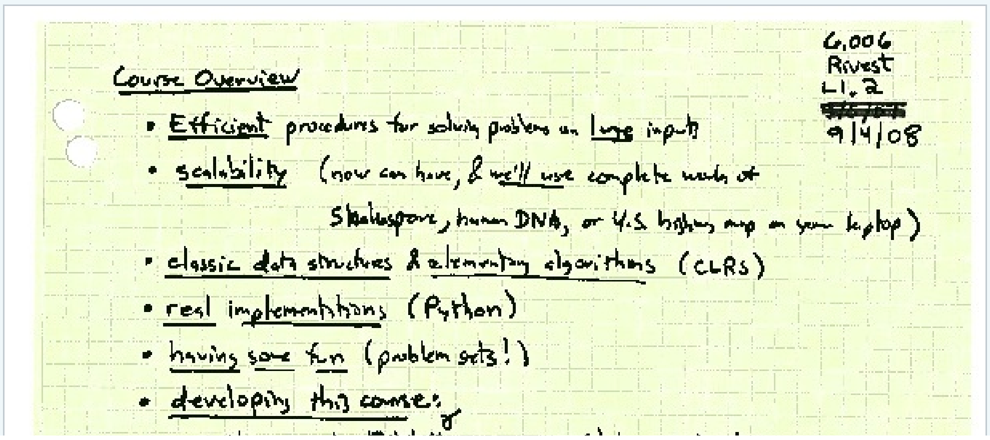
Outline:

- □ Administrivia
- □ Course overview
- □ "Document distance" problem

Handouts: 1. Course info
2. Sign-up sheet (passed around)
3. docdist1.py

Administrivia:

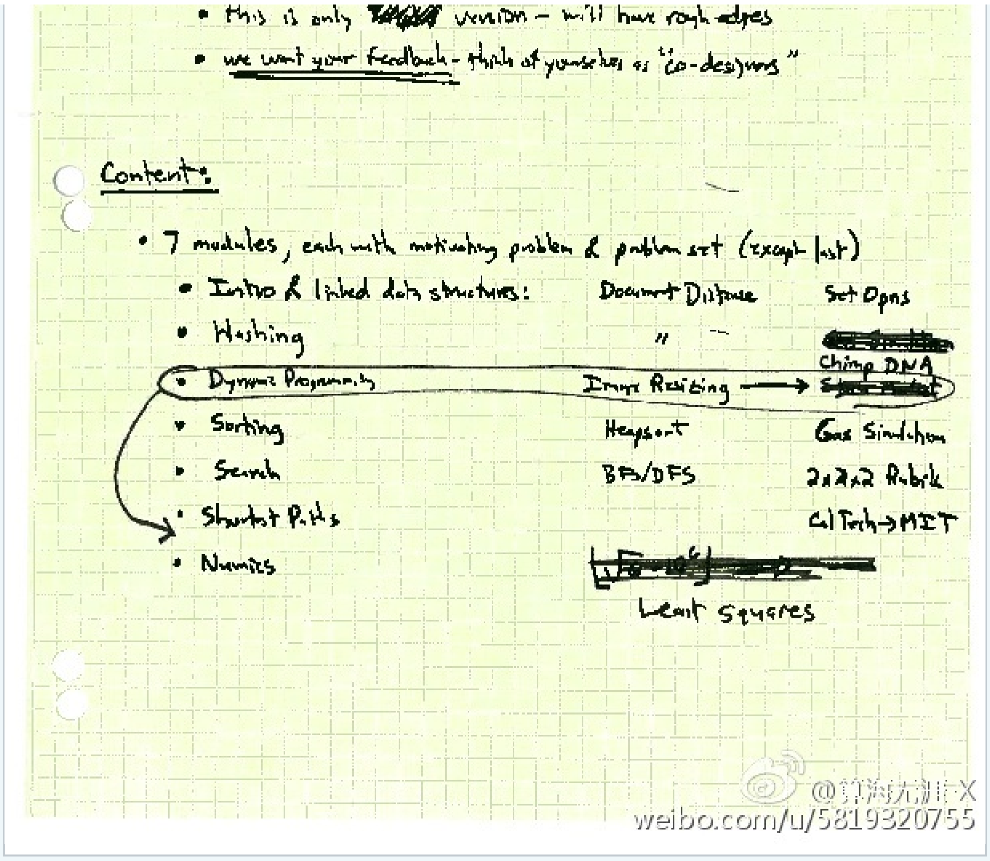
- Welcome to 6.006!
- Introduce staff
- Sign-up – on line
  and – on sheet being passed around
- web = http://courses.csail.mit.edu/6.006
- Ask students to raise hands for:
  credit/listener/unsure
  frosh/soph/...
  MIT/not
  course 6/18 other
  python
  6.01
  6.042
- Pre-reqs: see TA's if haven't got them, but have ≡
- Recitations: on-line; get assignment tonight – bring laptops!
- lectures/recitations/problemsets/2 quizes/final
- textbook: CLRS, rec: Miller/Ranum
- relation to 6.046



6.006
Rivest
L1.2
9/4/08

Course Overview

- Efficient procedures for solving problems on large inputs
- scalability (now can have, & we'll use complete works of
  Shakespeare, human DNA, or U.S. highway map on your laptop)
- classic data structures & elementary algorithms (CLRS)
- real implementations (Python)
- having some fun (problem sets!)
- developing this course:

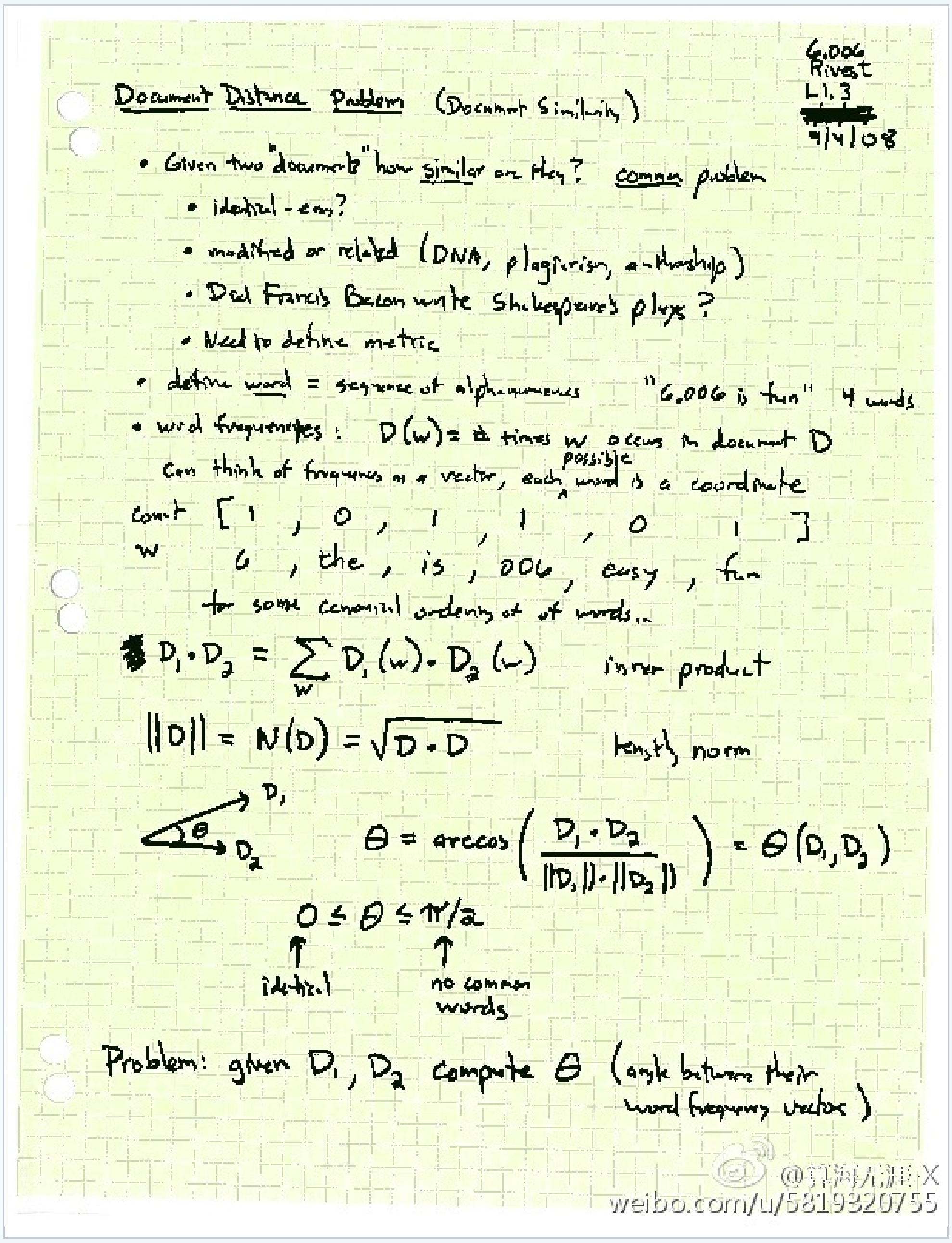


- this is only [redacted] version – will have rough edges
- we want your feedback – think of yourselves as "co-designers"

Content:

- 7 modules, each with motivating problem & problem set (except last)

• Intro & linked data structures:	Document Distance	Set Opns
• Hashing	"	[redacted]
• Dynamic Programming	Image Resizing →	Chimp DNA ~~[illegible]~~
• Sorting	Heapsort	Gas Simulation
• Search	BFS/DFS	2x2x2 Rubik
• Shortest Paths		CalTech→MIT
• Numerics	[redacted] Least squares	



6.006
Rivest
L1.3
[redacted]
9/4/08

Document Distance Problem (Document Similarity)

- Given two "documents" how similar are they? common problem
  - identical - easy?
  - modified or related (DNA, plagiarism, authorship)
  - Did Francis Bacon write Shakespeare's plays?
  - Need to define metric
- define word = sequence of alphanumerics "6.006 is fun" 4 words
- word frequencies: $D(w)$ = # times w occurs in document D

Can think of frequencies as a vector, each possible word is a coordinate

count [1 , 0 , 1 , 1 , 0 1]
w 6 , the , is , 006 , easy , fun

for some canonical ordering of of words..

$$D_1 \cdot D_2 = \sum_w D_1(w) \cdot D_2(w)$$ inner product

$$\|D\| = N(D) = \sqrt{D \cdot D}$$ length norm

D_1, D_2, θ

$$\theta = \arccos\left(\frac{D_1 \cdot D_2}{\|D_1\| \cdot \|D_2\|}\right) = \theta(D_1, D_2)$$

$$0 \le \theta \le \pi/2$$

↑ identical ↑ no common words

Problem: given D_1, D_2 compute θ (angle between their word frequency vectors)

2016年05月14日 00:24 来自 荣耀畅玩4X·快科技