

# CSCI 544 Fall 2018 course page

Jonathan May and Nanyun (Violet) Peng

August 20, 2018

Website	<a href="https://www.isi.edu/~jonmay/cs544_fa18_web/">https://www.isi.edu/~jonmay/cs544_fa18_web/</a>
Lectures	SAL 101, Wednesdays and Fridays 8:00-9:50 am
Instructor office hours	Jonathan May or Nanyun (Violet) Peng, RTH 512, Wednesdays and Fridays 10-11am or 11am-12pm with appointment
Textbook	None required; <i>Speech and Language Processing, 2nd Edition</i> <sup>1</sup> is optional but badly out of date. The new version (called “jm” in reading selection notes below) is incomplete but preferred <sup>2</sup> . All required readings from jm and elsewhere will be listed in the syllabus.
TA Office hours	Xusen Yin, 10:30am-12:30pm Mon., SAL Lab Ramesh Manuvinakurike, 11:30am-1:30pm Wed., SAL Lab Sarik Ghazarian, 3pm-5pm Wed., SAL Lab
Quizzes (3% of grade)	approximately 25 x 0.12% of grade each. Multiple choice online tests taken mid-class to break up the monotony. Open note, open internet, open discussion.
Homeworks (56.7% of grade)	7 x 8.1% of grade each. A mix of programming and written assignments, electronic submission due 11:59 PM local Pacific time per table below.
Late days	Four (4) <i>cumulative</i> with no more than two (2) per assignment. 50% penalty first day thereafter; no credit afterward.
Midterm (15% or 25% of grade, whichever is more advantageous)	Friday, October 5, 8:00-9:45 am. Covers lectures, readings, and homeworks from 8/22-10/3
Final exam (15% or 25% of grade, whichever is more advantageous)	Wednesday, December 5, 8:00-10:00 am. Covers lectures, readings, and homeworks from 8/22-11/30 (i.e. the entire class)
Contact us	On Piazza or in class/office hours. Please do not email (unless notified otherwise).

A note about the midterm/final grading: whichever you do better on will count as 25% of your grade and the other one will count as 15% of your grade. Together the tests will count as 40% of your grade.

# Syllabus/schedule – subject to change – latest version is always on the website!

instructor	date	material	reading	HW out	HW due
JM	8/22	intro, applications	jm v2 c1 <sup>3</sup> , hirschberg & manning <sup>4</sup>		
JM	8/24	corpora, text processing, words, regular expressions (lec 2 starting code <sup>5</sup> )	NLTK ch.2 <sup>6</sup> , jmv3 c2 <sup>7</sup> , nathan schneider's notes <sup>8</sup> , Unix for poets <sup>9</sup> , sculpting text <sup>10</sup> ,	HW 1 (due 9/7)	
JM	8/29	Morphology, Finite-State Automata, FSA relationship to regex, finite state transducers	jm v2 c3 <sup>11</sup>		
JM	8/31	probability	Goldwater probability tutorial <sup>12</sup> ,	HW 2 (due 9/14)	
JM	9/5	Classifiers, features, naive Bayes, perceptron, logistic regression	jm v3 c6 <sup>13</sup> , Eisenstein notes pp. 21–48 <sup>14</sup>		
JM	9/7	pos tagging, hmm, search	jm v3 c9 <sup>15</sup> , blunsom notes <sup>16</sup> , collins notes (optional, more detailed) <sup>17</sup>		HW1
NP	9/12	(guest lecture) parsing and syntax 1: treebanks, evaluation, cky, grammar induction	penn treebank <sup>18</sup> , jm v3 c12 <sup>19</sup> , jm v3 c13 <sup>20</sup>	HW3 (due 9/26)	
NP	9/14	(guest lecture) parsing and syntax 2: pcfgs, restructuring, lexicalization, smoothing, beaming	chiang notes <sup>21</sup>		HW2
NP	9/19	parsing and syntax 3: dependencies, mst and shift reduce algorithms	jm v3 c14 <sup>22</sup>	HW 4 (due 10/10)	
NP	9/21	ngram LMs	jm v3 c4 <sup>23</sup>		
NP	9/26	smoothing, interpolation	(optional) chen and goodman <sup>24</sup>		HW 3
NP	9/28	feed forward lm, rnn lm	jm v3 c8 <sup>25</sup>		
NP/JM (?)	10/3	REVIEW			
	10/5	MIDTERM			
NP	10/10	classical lexical semantics	jm v3 c17 <sup>26</sup>		HW 4
NP	10/12	distributional lexical semantics	jm v3 c15 <sup>27</sup> , jm v3 c16 <sup>28</sup>	HW 5 (due 10/26)	
NP	10/17	RNNs, LM and semantics	jmv3 c7 <sup>29</sup> c9 <sup>30</sup>		
NP	10/19	Information Extraction	jmv3 c21 <sup>31</sup>	HW 6 (due 11/9)	
NP	10/24	Information Extraction: CRF	Collins's Note <sup>32</sup> Sutton and McCallum's Note <sup>33</sup>		
NP	10/26	Neural CRF	Lample et. al. <sup>34</sup>		HW 5
JM	10/31	MT evaluation and alignment	arcturan and centauri <sup>35</sup> , Bleu <sup>36</sup> , TODO		
JM	11/2	Statistical MT	Brown et al. <sup>37</sup> (mathy bits of models 3+ can be skimmed), Knight workbook <sup>38</sup> , mert <sup>39</sup>		
JM	11/7	Neural MT	TODO		
JM	11/9	Summarization	TODO	HW 7 (due 11/30)	HW 6
JM	11/14	Generation, Dialogue	TODO		
JM	11/16	Dialogue	TODO		
	11/21	NO CLASS			
	11/23	NO CLASS			
JM	11/28	Applications	TODO		
JM/NP	11/30	Review			HW 7
	12/5	FINAL			