

Language & Technology

Syllabus

Thomas Graf

Stony Brook University
`lin120@thomasgraf.net`

Aug 28, 2016

Organizational Information

Course	Language and Technology
Course#	LIN 120
Room	Humanities 1003
Time	MW 2:30–3:23
Website	Blackboard

Instructor	Thomas Graf
Email	lin120@thomasgraf.net
Office hours	M 11:00–12:30 & 3:30–4:00 W 3:30–4:00 F 11:15–11:45
Office	SBS N249

TAs	Aniello de Santo & Alëna Aksënova
Undergraduate TAs	Orel Maimoni & Suji Yang

See the Blackboard course page for more details and updates.

An introduction to how computers process language and solve language-related tasks. This course discusses the language technologies of our daily life — spam filtering, machine translation, and many more — and shows how they work under the hood. The course explores a variety of issues: Why do computers do well in some areas (spell checking) yet fail miserably in others (essay grading)? Will we ever have perfectly fluent AIs as depicted in science fiction? And how will these technological advances impact the role of language in our society? Students will also acquire basic programming skills and write scripts for simple language tasks. No previous training in mathematics or computer science required.

SBC: TECH

3 credits

An Experiment

- 1 Open some chat or messaging app on your phone.
- 2 Don't type anything.
- 3 Instead, click the second word suggestion (the one in the middle).
- 4 Keep doing this.
- 5 Did you get a reasonable sentence of English?

An Experiment

- 1 Open some chat or messaging app on your phone.
- 2 Don't type anything.
- 3 Instead, click the second word suggestion (the one in the middle).
- 4 Keep doing this.
- 5 Did you get a reasonable sentence of English?

I am a beautiful person who is the best of luck to you by the way to get the best of luck to you by the way to get the best of luck to you by the way to get the . . .

**Current language technology is mostly
smoke and mirrors**

- ▶ How do computers process language?
- ▶ Why do they succeed in some areas (spell checker, spam filter), yet fail miserably in others (translation, poetry)?
- ▶ Will we ever have conversant AIs as depicted in science fiction (2001, Star Trek, Blade Runner, Her, Ex Machina, System Shock)?
- ▶ Can computers provide new answers to long-standing questions of linguistics and philology?
- ▶ How are language communities affected by these new technologies?

► **Basics of Programming and Computer Science**

- understand the importance of algorithms and data structures
- conceptualize linguistic problems in computational terms
- basic programming skills in Python

► **Cognitive Science**

- familiarity with notions of artificial intelligence
- understand how and why humans and computers differ in their linguistic abilities

► **Digital Humanities and Social Science**

- work with text corpora
- use computational tools for humanities (stylistic analysis, tracking social developments via corpora)
- understand the role of Big Data in computational linguistics
- awareness of the dangers of computational linguistics (surveillance, language death)

► **Basics of Programming and Computer Science**

- understand the importance of algorithms and data structures
- conceptualize linguistic problems in computational terms
- basic programming skills in Python

► **Cognitive Science**

- familiarity with notions of artificial intelligence
- understand how and why humans and computers differ in their linguistic abilities

► **Digital Humanities and Social Science**

- work with text corpora
- use computational tools for humanities (stylistic analysis, tracking social developments via corpora)
- understand the role of Big Data in computational linguistics
- awareness of the dangers of computational linguistics (surveillance, language death)

► **Basics of Programming and Computer Science**

- understand the importance of algorithms and data structures
- conceptualize linguistic problems in computational terms
- basic programming skills in Python

► **Cognitive Science**

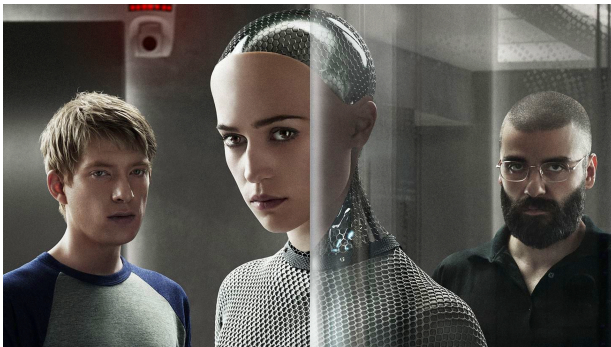
- familiarity with notions of artificial intelligence
- understand how and why humans and computers differ in their linguistic abilities

► **Digital Humanities and Social Science**

- work with text corpora
- use computational tools for humanities (stylistic analysis, tracking social developments via corpora)
- understand the role of Big Data in computational linguistics
- awareness of the dangers of computational linguistics (surveillance, language death)

Benchmark

By the end of the course, the following scene from *Ex Machina* should seem rather trivial to you.



▶ **What You Need**

- ▶ ability to operate a computer
(use a web browser, install software, edit text files)
- ▶ willingness to play around with open-ended problems

▶ **What You **WON'T** Need**

- ▶ programming experience
- ▶ math (except for addition, multiplication and fractions)
- ▶ linguistics (LIN 101 helps a bit, though)

Three Types of Instruction

Monday standard lecture on language technology

Wednesday programming sessions in Python

Recitation recap material with your TAs

Session	Mini-quiz?	Laptop?	Attendance?
Monday	yes	not recommended	recommended
Wednesday	no	recommended	recommended
Recitation	no	recommended	mandatory

Echo Video Recordings

- ▶ Video recordings of all lectures will be made available online.
- ▶ But the system is flaky, don't rely on it.

Three Types of Instruction

Monday standard lecture on language technology

Wednesday programming sessions in Python

Recitation recap material with your TAs

Session	Mini-quiz?	Laptop?	Attendance?
Monday	yes	not recommended	recommended
Wednesday	no	recommended	recommended
Recitation	no	recommended	mandatory

Echo Video Recordings

- ▶ Video recordings of all lectures will be made available online.
- ▶ But the system is flaky, don't rely on it.

Class Participation (20%)

- ▶ Both in class and **online!**
- ▶ **Examples:**
 - ▶ ask questions
 - ▶ help fellow students
 - ▶ link to relevant online materials
 - ▶ ⋮
- ▶ **Why?**
 - ▶ Encourages you to ask questions.
 - ▶ Helping others is a great way of learning.
 - ▶ I want to have some fun, too.

Low-Pressure Exercises (40%)

- ▶ once per week
- ▶ programming in **Python**
- ▶ assigned on Wednesday at 11:59pm
- ▶ due the following Wednesday at 11:59pm
- ▶ only pass-fail grading (0 points VS 1 point)
- ▶ no late hand-ins (more on that later)
- ▶ **Why?**
 - ▶ Learning programming is like learning a new language
⇒ needs constant practice
 - ▶ Even a little bit of programming experience is incredibly useful.

Mini-Quizzes (40%)

- ▶ at beginning of Monday lectures
- ▶ apply techniques discussed in lecture
converting number to binary, listing all character trigrams of a word, . . .
- ▶ question templates in online **quiz pool**
- ▶ only pass-fail grading (0 points VS 1 point)
- ▶ missed quiz is an automatic fail (more on that later)
- ▶ **Why?**
 - ▶ We want you to learn skills and techniques, not definitions.
 - ▶ Quizzes force you to self-assess how much you're getting out of the class.

- ▶ Approximately every 4 weeks you can hand in optional **skill assessments**.
 - ▶ 1 for Python
 - ▶ 1 for Theory
- ▶ These are take-home exams with multiple assignments.
- ▶ Each successfully completed assignment converts 1 Fail into a Pass for the respective track (Python/Theory).

Example

Our example student Stu has the following grades after seven weeks:

	1	2	3	4		5	6	7	
Quiz	P	P	F	P	3/4	F	F	P	0/3
Python	P	P	F	F	1/4	F	P	F	2/3

Diagram illustrating the relationship between the assessments and the skill assessment:

```
graph TD; A[3/4] --- B[1/4]; A --- C[Skill assessment]; B --- C; C --- D[0/3]; C --- E[2/3];
```

- 1 His first theory assessment converts quiz 3 from F to P.
- 2 Quizzes 5 & 6 remain F because he got 0 points on assessment 2. Points from other assessments cannot be used!
- 3 His first Python assessment converts exercise 3 from F to P.
- 4 His second Python assessment converts exercise 5 and 7 to P.

Example

Our example student Stu has the following grades after seven weeks:

	1	2	3	4		5	6	7	
Quiz	P	P	P	P	3/4	F	F	P	0/3
Python	P	P	F	F	1/4	F	P	F	2/3

Diagram illustrating the skill assessment process:

- A box labeled "3/4" (Quiz score) is connected by a line to a box labeled "1/4" (Python score).
- A box labeled "0/3" (Quiz score) is connected by a line to a box labeled "2/3" (Python score).
- Both "1/4" and "2/3" boxes are connected by lines to a central box labeled "Skill assessment".

- 1 His first theory assessment converts quiz 3 from F to P.
- 2 Quizzes 5 & 6 remain F because he got 0 points on assessment 2. Points from other assessments cannot be used!
- 3 His first Python assessment converts exercise 3 from F to P.
- 4 His second Python assessment converts exercise 5 and 7 to P.

Example

Our example student Stu has the following grades after seven weeks:

	1	2	3	4		5	6	7	
Quiz	P	P	P	P	3/4	F	F	P	0/3
Python	P	P	F	F	1/4	F	P	F	2/3

Diagram illustrating the relationship between assessments and skill assessments:

- A box labeled "3/4" (Quiz 4) is connected by a line to a box labeled "1/4" (Python 4).
- A box labeled "0/3" (Quiz 7) is connected by a line to a box labeled "2/3" (Python 7).
- Both "1/4" and "2/3" boxes are connected by lines to a central box labeled "Skill assessment".

- 1 His first theory assessment converts quiz 3 from F to P.
- 2 Quizzes 5 & 6 remain F because he got 0 points on assessment 2. Points from other assessments cannot be used!
- 3 His first Python assessment converts exercise 3 from F to P.
- 4 His second Python assessment converts exercise 5 and 7 to P.

Example

Our example student Stu has the following grades after seven weeks:

	1	2	3	4		5	6	7	
Quiz	P	P	P	P	3/4	F	F	P	0/3
Python	P	P	P	F	1/4	F	P	F	2/3

Diagram illustrating the relationship between assessments and skill assessments:

- A box labeled "3/4" (Quiz 4) is connected by a line to a box labeled "1/4" (Python 4).
- A box labeled "0/3" (Quiz 7) is connected by a line to a box labeled "2/3" (Python 7).
- Both "1/4" and "2/3" boxes are connected by lines to a central box labeled "Skill assessment".

- 1 His first theory assessment converts quiz 3 from F to P.
- 2 Quizzes 5 & 6 remain F because he got 0 points on assessment 2. Points from other assessments cannot be used!
- 3 His first Python assessment converts exercise 3 from F to P.
- 4 His second Python assessment converts exercise 5 and 7 to P.

Example

Our example student Stu has the following grades after seven weeks:

	1	2	3	4		5	6	7	
Quiz	P	P	P	P	3/4	F	F	P	0/3
Python	P	P	P	F	1/4	P	P	P	2/3

Diagram illustrating the relationship between assessments and skill assessments:

- A box labeled "3/4" (Quiz 4) is connected by a line to a box labeled "1/4" (Python 4).
- A box labeled "0/3" (Quiz 7) is connected by a line to a box labeled "2/3" (Python 7).
- Both "1/4" and "2/3" are connected by lines to a central box labeled "Skill assessment".

- 1 His first theory assessment converts quiz 3 from F to P.
- 2 Quizzes 5 & 6 remain F because he got 0 points on assessment 2. Points from other assessments cannot be used!
- 3 His first Python assessment converts exercise 3 from F to P.
- 4 His second Python assessment converts exercise 5 and 7 to P.

Soapbox: My Thoughts on Grades

- ▶ Students are caught up in the **grade bubble**:
 - ▶ If I get good grades I will get a job.
 - ▶ If I get bad grades I will fail in life.
- ▶ This is wrong, wrong, wrong, wrong!
- ▶ In the real world, nobody cares about your GPA.
- ▶ Don't focus on grades!
- ▶ Focus on mastering the skills you need to get the job you want.

Soapbox: My Role in This

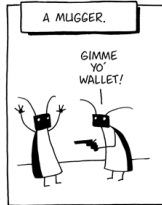
- ▶ I am the academic equivalent of a **fitness trainer**.
- ▶ You're paying thousands of dollars for me to get you into shape, and I've developed a program for you that will do that.
- ▶ But you are the one who has to move their body.
- ▶ Bad techniques like cram learning may get you a good grade, but you're cheating yourself out of true progress.
- ▶ If you aren't working towards long-term intellectual growth, you're flushing tons of money down the toilet.

Soapbox: A Rough Cost Calculation

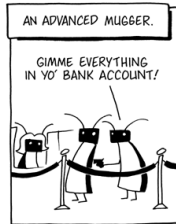


© 2009 ADAM HUBER

Soapbox: A Rough Cost Calculation



© 2009 ADAM HUBER



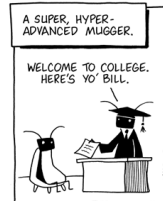
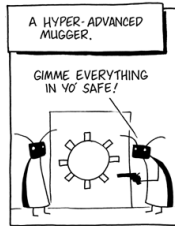
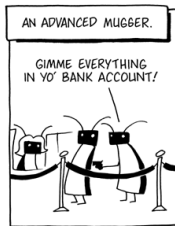
Soapbox: A Rough Cost Calculation



Soapbox: A Rough Cost Calculation



© 2009 ADAM HUBER

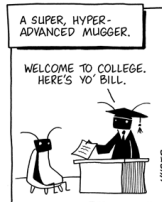
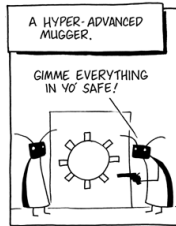
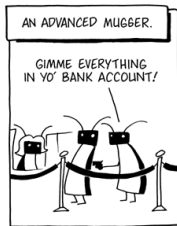


WWW.BUGCOMIC.COM

Soapbox: A Rough Cost Calculation



© 2009 ADAM HUBER



WWW.BUGCOMIC.COM

Expense	Resident	Non-Resident
Tuition & Fees, 12cr	\$5,876.25	\$14,631.25
Housing & Meals	~\$4,500.00	\$4,500.00
Cost of LIN 120	\$2594.06	\$4782.81
Cost per week	\$185.29	\$341.63

- ▶ By default: use discussion forums on Blackboard
- ▶ Contacting me:
 - ▶ `lin120@thomasgraf.net`
 - ▶ Not using this address means delayed replies!
 - ▶ Reply time usually < 24h
 - ▶ If you plan to come to my office hours, please drop me a line the day before. If there's a scheduling conflict, I'll let you know. Radio silence means everything is fine.
- ▶ For additional instructions, see the *Getting Help* section on Blackboard.

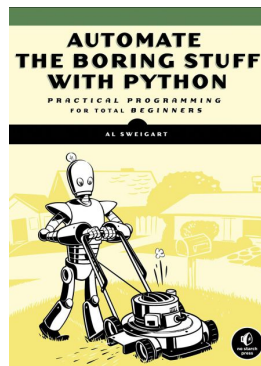
1 Course Website

- ▶ Familiarize yourself with the course website.
- ▶ Lots of extra information there.

2 Software Setup

- ▶ We will be using **Jupyter notebooks**.
- ▶ See the installation instructions on Blackboard.
- ▶ More information in Wednesday lecture.

- ▶ Al Sweigart (2015):
Automate the Boring Stuff with Python
- ▶ online version free
- ▶ digital versions and hardcopy around \$25
- ▶ supplementary videos on Youtube
- ▶ Only a few parts will be mandatory, but it's a good source to consult if something is unclear.



Your First Homework

- 1 Carefully re-read this syllabus.
- 2 Read the document *How to Ace This Class* (it's on Blackboard).
- 3 There will be a mini-quiz on Wednesday to test your understanding of the course modalities (bring pen and paper!).

Hint: Check the quiz pool online for example questions.

Disability Support Services

If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Disability Support Services, ECC (Educational Communications Center) Building, Room 128, (631) 632-6748. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and Disability Support Services. For procedures and information go to the following website:

<http://www.stonybrook.edu/ehs/fire/disabilities>

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at <http://www.stonybrook.edu/uaa/academicjudiciary/>

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Judicial Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures.