

Natural Language Processing

113

Brief History of NLP

The Turing Test

- Alan Turing: the *Turing test*
 - language as test for intelligence
- Three participants
 - a computer and two humans (one is an interrogator)
- Interrogator's goal
 - to tell the machine and human apart
- Machine's goal
 - to fool the interrogator into believing that a person is responding
- Other human's goal
 - to help the interrogator reach his goal

Q: Please write me a sonnet on the topic of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: 105621 (after a pause)

Eliza

User₁: You are like my father in some ways.

ELIZA₁: WHAT RESEMBLANCE DO YOU SEE

User₂: You are not very aggressive but I think you don't want me to notice that.

ELIZA₂: WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

User₃: You don't argue with me.

ELIZA₃: WHY DO YOU THINK I DON'T ARGUE WITH YOU

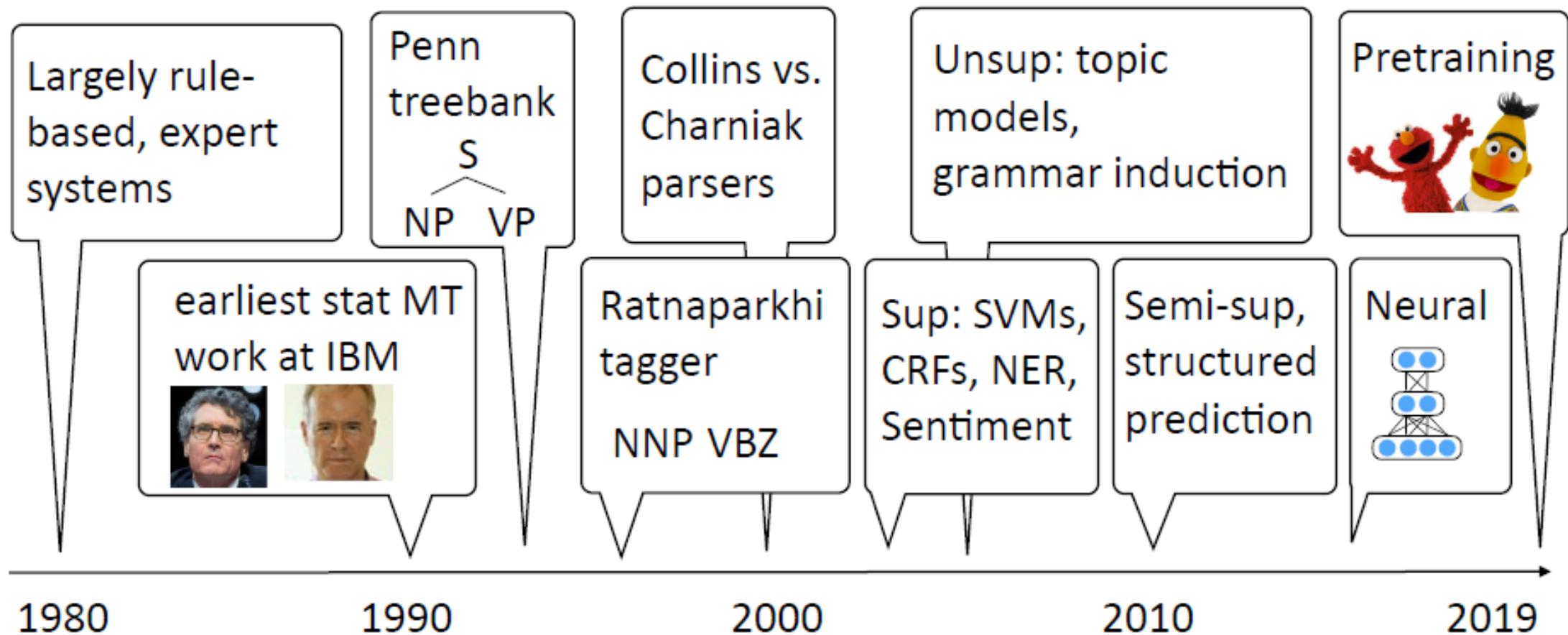
User₄: You are afraid of me.

ELIZA₄: DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

Longer example here:

<http://thoughtcatalog.com/oliver-miller/2012/08/a-conversation-with-eliza>

A Brief History of NLP



[Slide from Greg Durrett]

Some Brief History

- Foundational insights (1940's and 1950's)
 - automaton (Turing)
 - probabilities
 - information theory (Shannon)
 - formal languages (Backus and Naur)
 - noisy channel and decoding (Shannon)
 - first systems (Davis et al., Bell Labs)
- Two camps (1957-1970): symbolic and stochastic
 - Transformation Grammar (Harris, Chomsky)
 - Artificial Intelligence (Minsky, McCarthy, Shannon, Rochester)
 - automated theorem proving and problem solving (Newell and Simon)
 - Bayesian reasoning (Mosteller and Wallace)
 - Corpus work (Kučera and Francis)

Some Brief History

- Four paradigms (1970-1983)
 - stochastic (IBM)
 - logic-based (Colmerauer, Pereira and Warren, Kay, Bresnan)
 - nlu (Winograd, Schank, Fillmore)
 - discourse modelling (Grosz and Sidner)
- Empiricism and finite-state models redux (83-93)
 - Kaplan and Kay (phonology and morphology)
 - Church (syntax)
- Late years (1994-2010)
 - integration of different techniques
 - different areas (including speech and IR)
 - probabilistic models
 - machine learning
 - structured prediction
 - topic models

The Most Recent Years

- Machine learning methods
 - SVM, logistic regression (maxent), CRFs
- Shared tasks
 - TREC, DUC, TAC, *SEM
- Semantic tasks
 - RTE, SRL
- Semi-supervised and unsupervised methods
 - Zero-shot learning, transfer learning, self training
- Deep Learning
 - Embeddings, LSTM, CNN, Attention, GANs, Transformers, BERT

Natural Language Processing

111

Introduction

NLP Applications

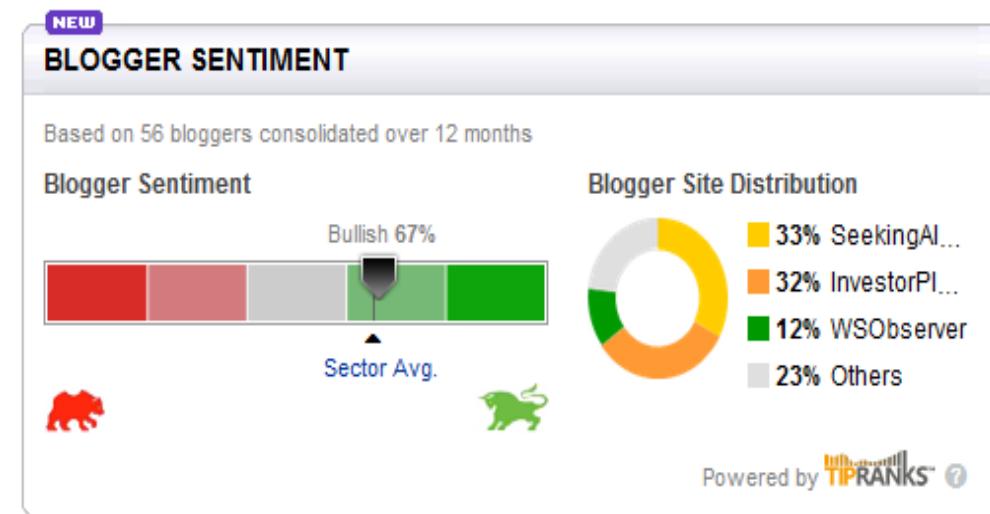
Text Documents

Detect Language English Bulgarian Albanian French Bulgarian English

California-grown avocados sold in bulk to retail stores in six states by the Henry Avocado Corporation are being recalled due to potential contamination with the bacterium *listeria monocytogenes*, the company announced on Saturday. "Henry Avocado is issuing this voluntary recall out of an abundance of caution due to positive test results on environmental samples taken during a routine government inspection at its California packing facility," reads the company's statement. There have been no reported illnesses associated with the recall at this time. The recalled products included California-grown conventional and organic avocados. They were packed at Henry Avocado Corporation's facility in California and distributed to six states: Arizona, California, Florida, New Hampshire, North Carolina and Wisconsin. Henry Avocado, a family-owned and managed company, started packing in this facility in late January 2019 and every shipment is subject to the recall.

Les avocats de Californie vendus en vrac à Henry Avocado Corporation dans des magasins de vente au détail dans six États sont rappelés en raison d'un risque de contamination par la bactérie *listeria monocytogenes*, a annoncé la société samedi. "Henry Avocado lance ce rappel volontaire par prudence, en raison des résultats de tests positifs sur des échantillons environnementaux prélevés au cours d'une inspection de routine effectuée par le gouvernement dans son usine d'emballage en Californie", lit-on dans le communiqué. Aucun cas de maladie associé au rappel n'a été signalé à ce jour. Les produits rappelés comprenaient des avocats conventionnels et biologiques cultivés en Californie. Ils ont été emballés dans les installations de Henry Avocado Corporation en Californie et distribués dans six États: Arizona, Californie, Floride, New Hampshire, Caroline du Nord et Wisconsin. Henry Avocado, une entreprise familiale, a commencé à emballer ses produits dans cette installation à la fin du mois de janvier 2019 et chaque envoi est sujet au rappel.

965/5000



NLP Applications

- Search engines
 - Google, Bing, DuckDuckGo, Baidu, Yandex
- Question answering
 - IBM Watson
- Natural language assistants
 - Siri, Alexa, Google, Cortana
- Machine translation
 - Google Translate
- Email assistants
 - Autocomplete

What is Natural Language Processing

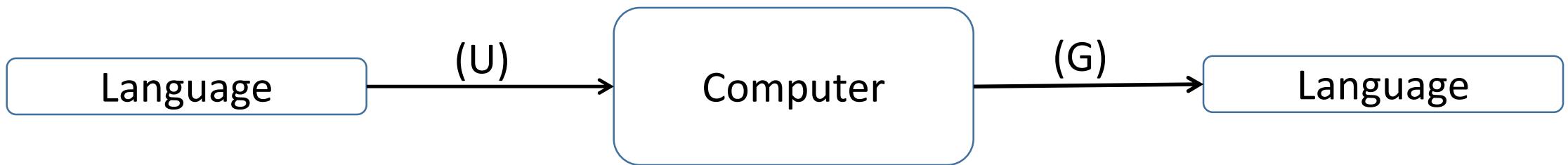
- **Natural Language Processing (NLP)**
 - The study of the computational treatment of natural (human) language.
 - In other words, building computers that understand (and generate) language.
- **Computational Linguistics (CL)**
 - The use of computers to study language
 - Applications to social science, literature, finance
- Computers are designed to deal with human language
 - Specific techniques are needed
- NLP draws on research in many fields
 - Linguistics, Theoretical Computer Science, Mathematics, Statistics, Artificial Intelligence, Psychology, Databases, etc.

Language and Communication

- Speaker
 - Intention (goals, shared knowledge and beliefs)
 - Generation (tactical)
 - Synthesis (text or speech)
- Listener
 - Perception
 - Interpretation (syntactic, semantic, pragmatic)
 - Incorporation (internalization, understanding)
- Both
 - Context (grounding)

Basic NLP Pipeline

- (U)nderstanding and (G)eneration



Representation Matters

Representation:
The Earth is fixed center of
our Solar System

Representation:
The Sun is fixed center of
our Solar System



Geocentric Model

(Anaximander, 6th century BC)

Heliocentric Model

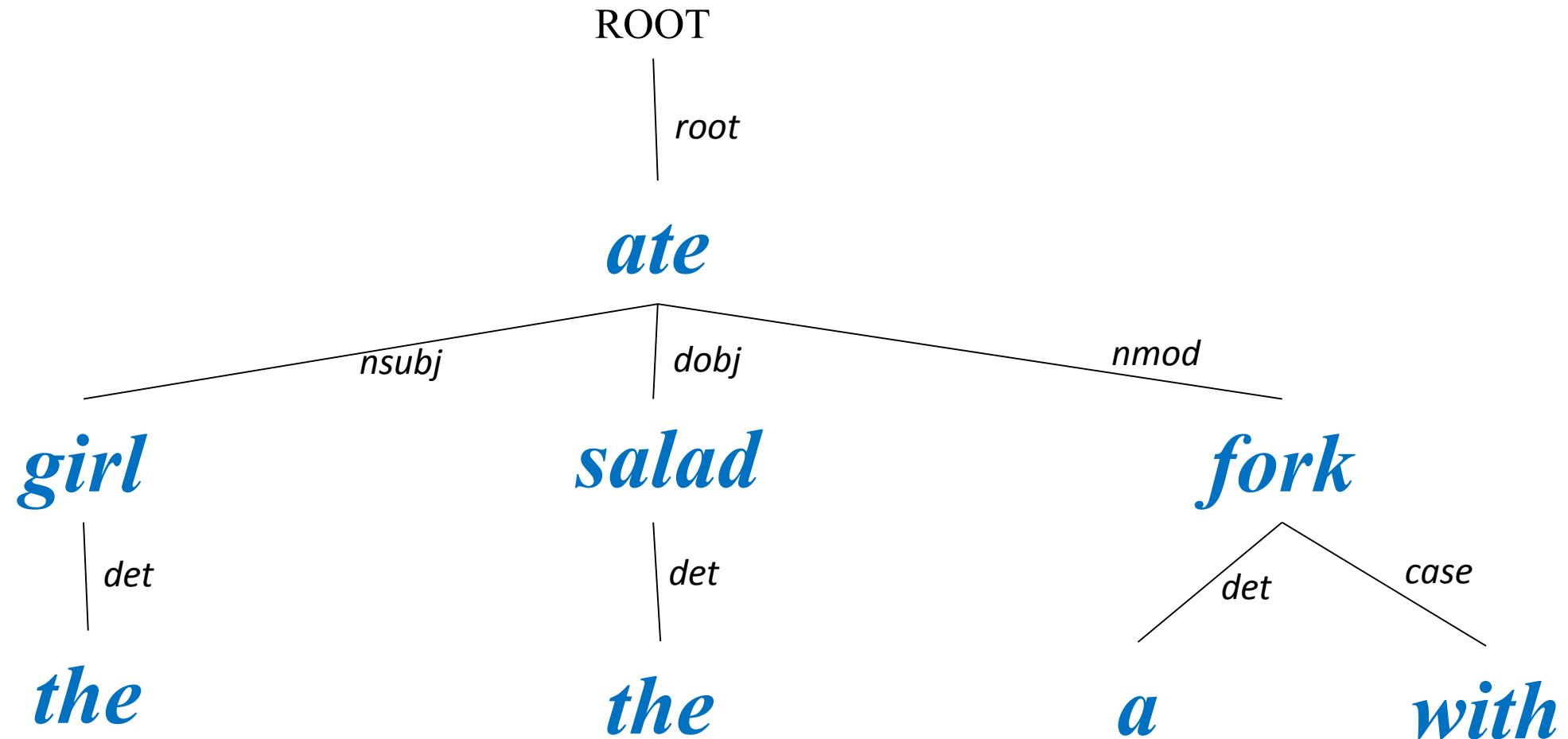
(Copernicus, 1543)

Natural Language Understanding

DET N VBD DET N PRP DET N

The girl ate the salad with a fork.

Sentence Representation



Logical Language Understanding

- Semantic Analysis

Girl (g_1)

Salad (s_1)

Fork (f_1)

EatingEvent (e_1) \wedge Eater (e_1, g_1) \wedge Eaten (e_1, p_1) \wedge Instrument (e_1, f_1) \wedge Time(e_1 , "1:24 pm")

- World Knowledge

$\forall x: \text{Salad}(x) \Rightarrow \text{Food}(x)$

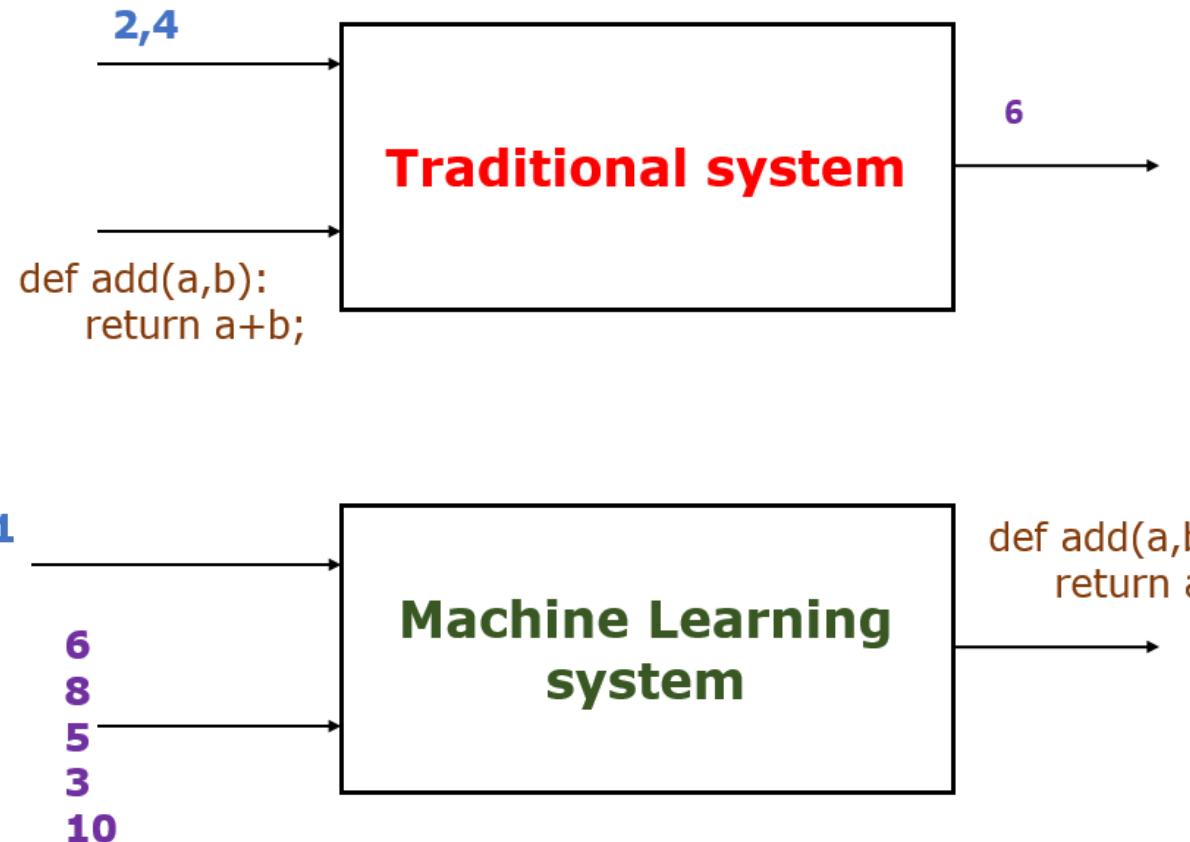
- Inference

$\forall z, t_0, t_1, y, e: \text{Hungry}(z, t_0) \wedge \text{EatingEvent}(e) \wedge \text{Eater}(e, z) \wedge \text{Eaten}(e, y) \wedge \text{Time}(e, t_0)$
 $\wedge \text{Food}(y) \wedge \text{Precedes}(t_0, t_1) \Rightarrow \neg \text{Hungry}(z, t_1)$

- Conclusion

$\neg \text{Hungry}(g_1, \text{now})$

Machine Learning



Modern NLP Techniques (2013-2021)

- Deep Learning
- Word Embeddings
- Convolutional Neural Networks
- Recurrent Neural Networks
- Attention
- Transformers
- BERT

Distributed representations

- acerola is a significant source of vitamin C.
- the pulp of the acerola is very soft
- acerola are now found growing in most sub-tropical regions of the world.
- acerola can be eaten fresh or used to make jams or jellies.

Word Embeddings

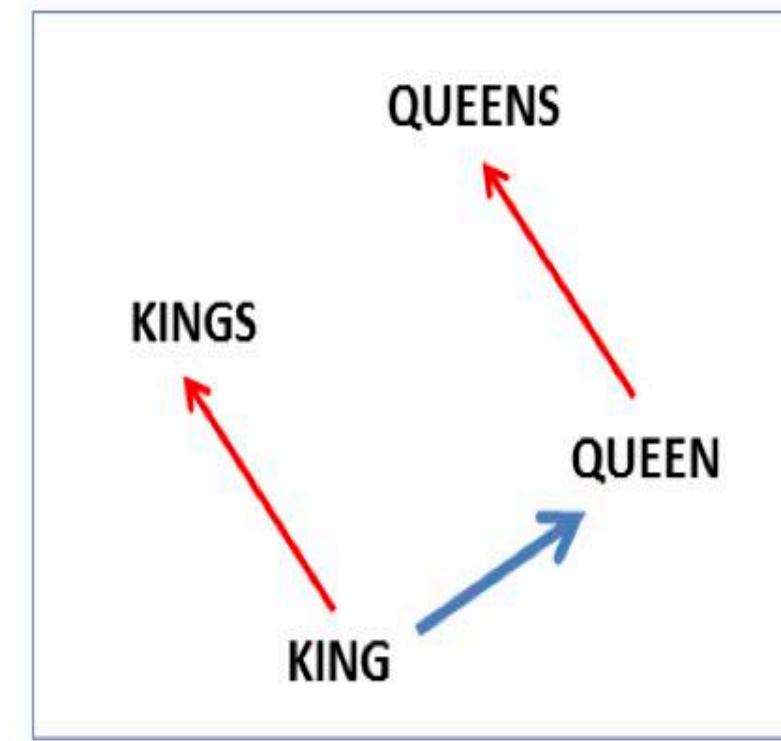
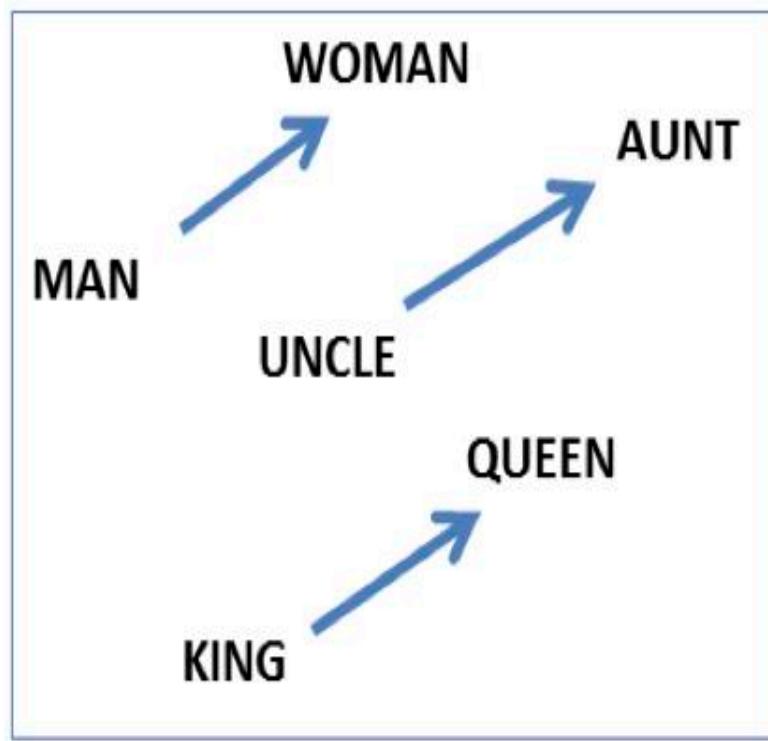
pharmaceuticals

```
[-0.43337 0.10411 -0.49926 -0.06442 0.39005 0.65823 -1.3033 -0.39019 -0.41713 -0.038667 -0.37677 0.35303 -0.16425 -0.46671 -
0.24358 0.10142 0.54664 0.065344 0.022733 -0.10422 -0.52839 0.20096 -0.23819 0.28856 0.098377 -0.12174 -0.035851 -0.92717
0.28408 0.22439 0.027598 0.47919 -0.51142 -0.062696 0.13002 0.20366 -0.050135 -0.4593 0.63779 -0.56294 -0.40989 -0.60342 0.8151
0.024929 -0.0096054 0.55748 0.35156 0.14476 -0.36327 -0.34856 0.1787 0.76398 0.27803 0.11999 0.42814 0.17844 0.092123 0.058811
-0.41114 -0.12101 -0.34001 -0.49946 -0.073549 0.15371 0.18034 0.34176 0.072738 -0.23442 -0.023682 -0.2499 -0.17334 -0.15008 -
0.14599 0.51706 0.52797 -0.075328 0.13018 -0.069498 -0.15378 -0.31615 0.59434 -0.91396 -0.12803 0.32963 0.70337 -0.095882 -
0.37066 0.16993 -0.62115 -0.76234 0.49005 -0.026823 -0.35171 -0.070227 0.19778 0.25563 -1.4504 -0.47122 -0.10107 -0.18279 -
0.31553 0.090524 0.1975 0.073745 0.34809 -0.26728 -0.04808 -0.18467 0.18147 0.37255 0.26197 -0.0046708 0.51 -0.99408 -0.1942 -
0.82518 0.59211 0.31112 0.3472 -0.066567 0.65975 -0.52254 -0.48302 0.30366 -0.35524 0.0022488 -0.89521 -0.096487 -0.36811 -
0.1139 -0.039127 0.03701 0.18691 -0.28874 0.19926 -0.71229 0.21108 0.1768 0.27541 -0.72828 -0.74097 0.15007 -0.46696 0.52759
0.69806 0.28434 1.2781 0.033105 0.2153 -0.59069 -0.18089 -0.28775 -0.30792 -0.32764 -0.20838 -0.49774 0.2604 -0.26116 -0.29203 -
0.18311 0.016024 0.26013 -0.4441 0.11857 0.61598 -0.25685 -0.49715 0.58277 -0.093157 -0.078187 -0.18587 -0.021307 0.52742
0.75704 0.091185 -0.41006 -0.26896 0.43715 0.13183 -0.49085 -0.84639 -0.22379 -0.094786 0.35858 -1.16 -0.019064 -0.29052 0.21588
0.026218 0.22063 -0.64061 0.89117 -0.14541 -0.47563 0.77044 -0.45668 0.49585 0.45303 -0.24904 0.24502 0.42608 0.0077214 -
0.55742 0.17449 -0.2142 0.26996 -0.26239 0.18933 -0.66798 -0.004951 0.062785 0.45616 -0.77372 0.29266 -0.76515 0.2079 -0.52916 -
0.13621 -0.60588 -0.049171 -0.21234 -0.071004 0.092045 0.87973 -1.0929 -0.028515 -0.28424 0.26105 0.2524 0.35996 -0.74328 -
0.27066 -0.046789 0.081494 0.70996 0.18443 -0.10652 -0.32352 -0.026315 0.079707 0.14203 -0.21906 0.49254 -0.29718 0.25746 -1.108
-0.35566 -0.14188 -0.35727 -0.34243 -0.37111 1.0034 -0.13679 0.92309 -0.16716 -0.36536 0.019904 -0.33768 0.18646 0.4391 0.045023
-0.25726 -0.14073 0.77022 -0.18926 -0.27791 0.2978 -0.78997 0.052217 0.72518 0.0089245 0.1029 1.0213 -0.70057 0.31295 0.29313 -
0.53556 0.75132 0.29351 -0.70482 -0.61882 -0.33732 0.60293 -0.33575 -0.12536 0.27659 -0.20361 0.12683 0.10469 -0.47956 0.187
0.38118 0.16238 -0.0484 0.43112 0.0089624 0.0051162 -0.67922 0.1709 -0.020472]
```

[Glove 300d vector]

Semantic Compositionality

- $\text{vector}(\text{'king'}) - \text{vector}(\text{'man'}) + \text{vector}(\text{'woman'}) \approx \text{vector}(\text{'queen'})$



[Mikolov 2013]

Epoch
000,000Learning rate
0.03Activation
SigmoidRegularization
NoneRegularization rate
0Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 80%

Noise: 0

Batch size: 5

REGENERATE

FEATURES

Which properties do you want to feed in?

 x_1
 x_2
 x_1^2
 x_2^2
 x_1x_2
 $\sin(x_1)$
 $\sin(x_2)$

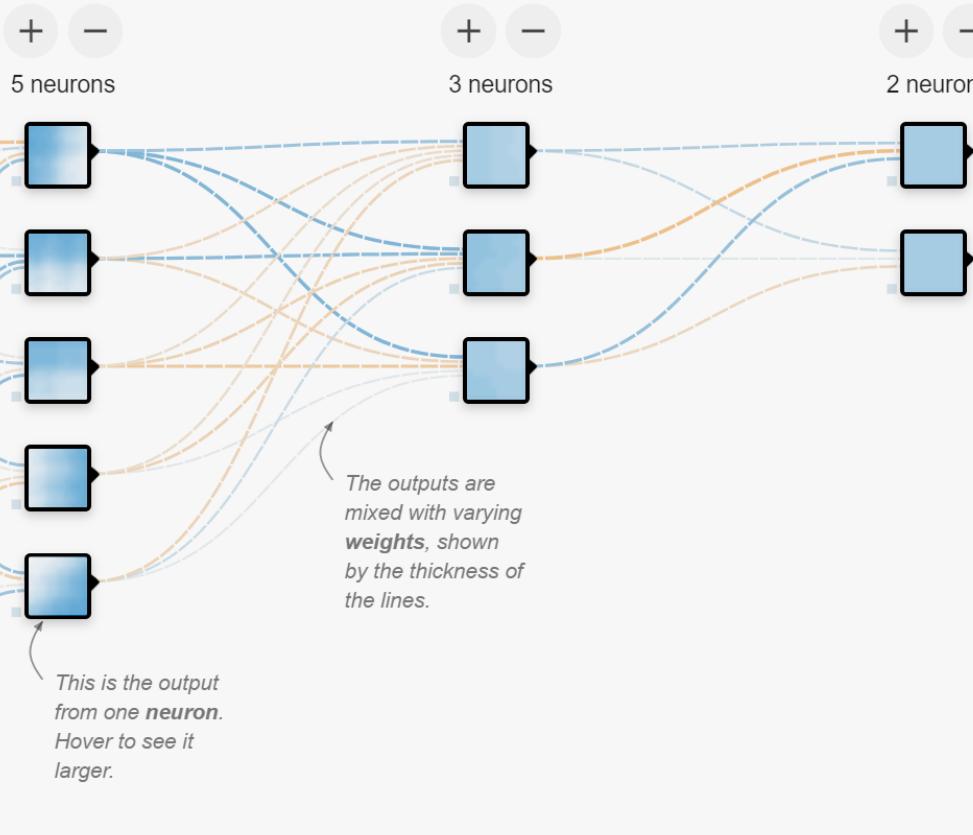
+ - 3 HIDDEN LAYERS

+ -

5 neurons

3 neurons

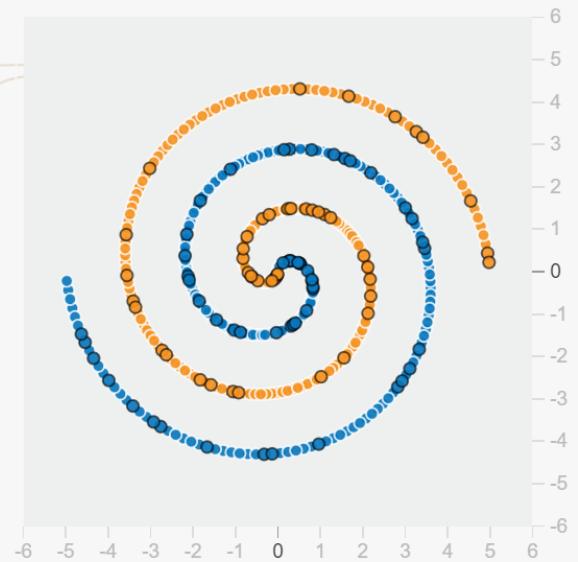
2 neurons



OUTPUT

Test loss 0.501

Training loss 0.500

 Show test data Discretize output

Epoch
000,785Learning rate
0.03Activation
SigmoidRegularization
NoneRegularization rate
0Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 80%

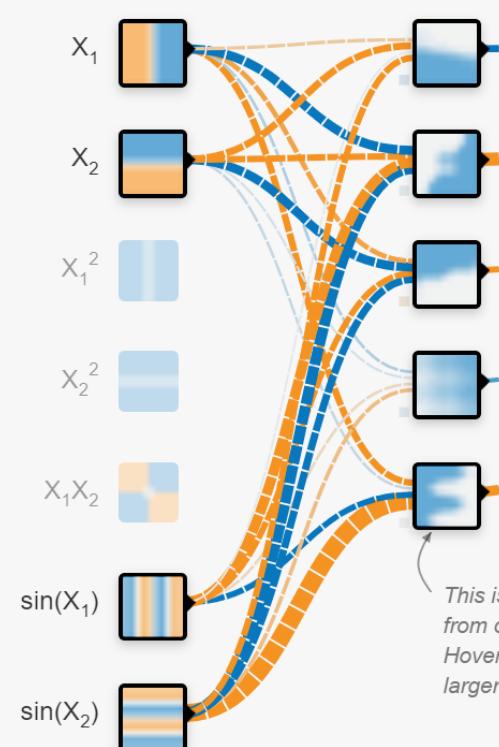
Noise: 0

Batch size: 5

REGENERATE

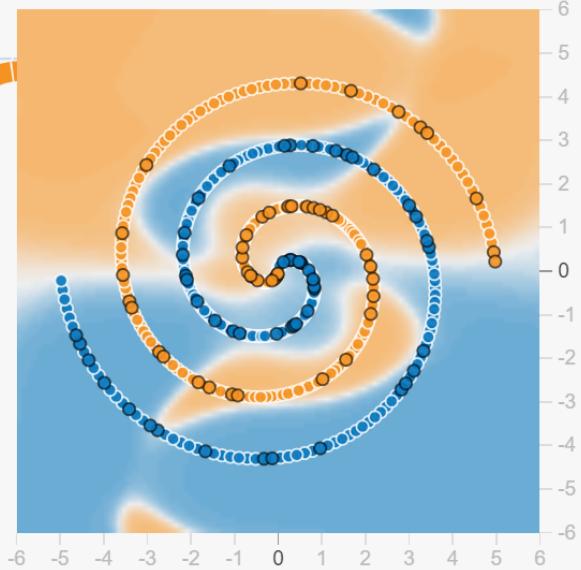
FEATURES

Which properties do you want to feed in?



3 HIDDEN LAYERS

OUTPUT

Test loss 0.272
Training loss 0.217 Show test data Discretize output

Epoch
023,073Learning rate
0.03Activation
SigmoidRegularization
NoneRegularization rate
0Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 80%

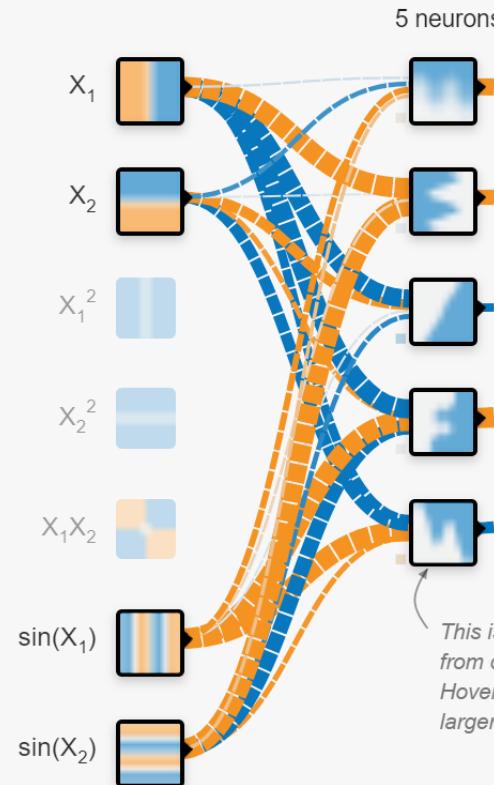
Noise: 0

Batch size: 5

REGENERATE

FEATURES

Which properties do you want to feed in?



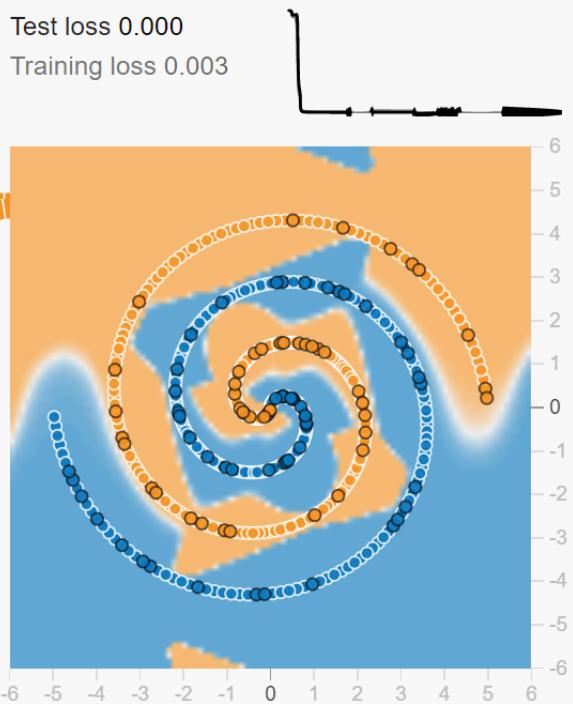
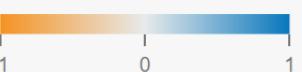
+ - 3 HIDDEN LAYERS

+ -

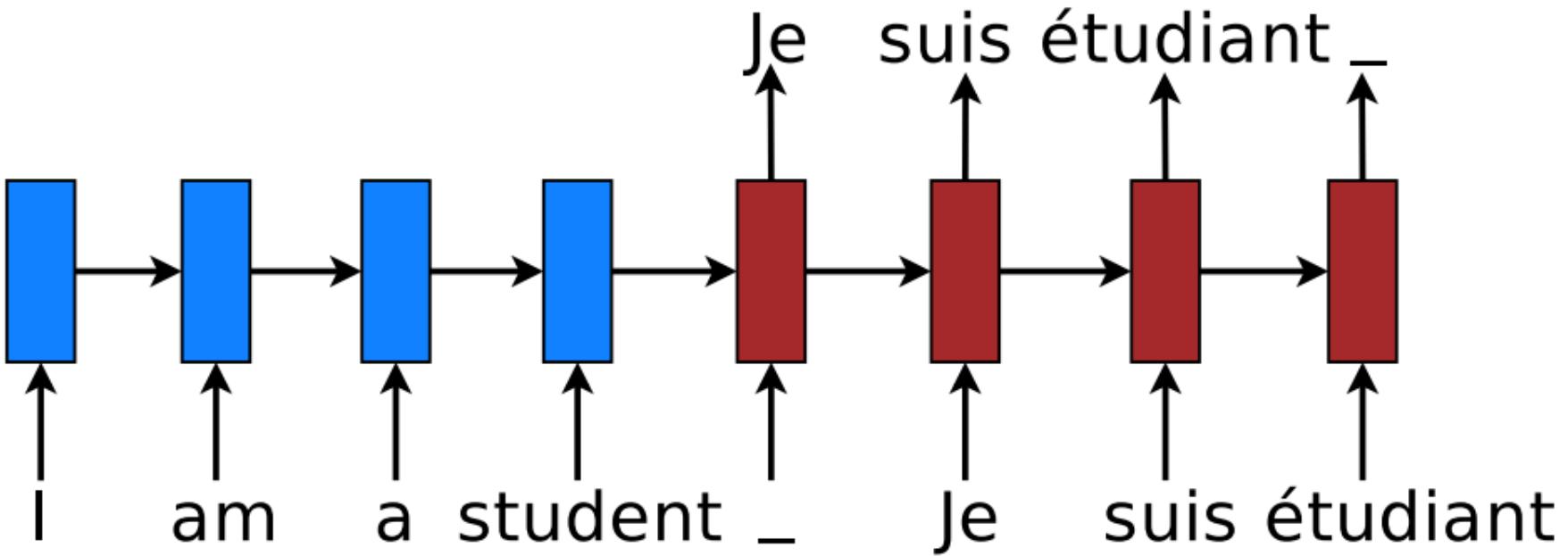
+ -

+ -

OUTPUT

Test loss 0.000
Training loss 0.003Colors shows
data, neuron and
weight values. Show test data Discretize output

Sequence to Sequence



[Sutskever, Vinyals, and Le, 2014]

Natural Language Processing

112

Examples of Text

What NLP is not about

- Romeo loves Juliet.
- ZZZZZ is a great stock to buy.

What it is about (1/2)

- After the ball, in what is now called the "balcony scene", Romeo sneaks into the Capulet orchard and overhears Juliet at her window vowing her love to him in spite of her family's hatred of the Montagues.

What it is about (2/2)

- **ZZZZZ Resources** (NYSE:ZZZZZ) in their third quarter financials present a picture of a company with a relatively high amount of debt versus shareholder equity, and versus revenues. The company had total liabilities in the third quarter of \$4,416 million versus shareholders' equity of only \$1,518 million. That is a very high 3 to 1 debt to equity ratio. The company had third quarter revenues of \$306 million. On an annualized basis, revenues would come out to \$1,224 million. The company's debt level is almost 3 times its annual revenues. And remember, third quarter revenue is from before oil prices dropped in half. It looks like ZZZZZ may have bitten off more than it can chew.
- **XXXXX Petroleum** (NYSE:XXXXX) is another company whose third quarter financials present a relatively high debt load. The company had total liabilities in the third quarter of \$3,272 million versus shareholder equity of only \$1,520 million. That represents a high 2 to 1 debt to equity ratio. The company had third quarter revenues of \$350 million. On an annualized basis revenues would come out to \$1,400 million. The company's debt is more than 2 times its annual revenue. While XXXXX is a very good operator, it looks like they have taken on the high debt strategy at the wrong time.
- **YYYYY Energy** (NYSE:YYYYY) has a relatively high debt load according to their third quarter financials. The company had total liabilities of \$2,026 million versus shareholder equity of \$1,079. That is almost a 2 to 1 debt to equity ratio. Their third quarter revenues were \$207 million. When annualized, their third quarter revenues come out to \$827 million. The company's debt is almost 2 1/2 times its annualized revenues, and that is before the collapse of oil prices in the fourth quarter. YYYYY has taken the Brigham model to heart and has been aggressively growing the company.

Genres of Text

- Blogs, emails, press releases, chats, debates, etc.
- Each presents different challenges to NLP

The screenshot shows a web browser window for 'CreateDebate'. The main content area displays a debate titled 'Do you think that schools should provide more help to developmentally challenged'. Two main arguments are shown: 'yes they should' (Side Score: 1) and 'what they are doing is fine' (Side Score: 2). Below each argument, there are several user comments and replies. A sidebar on the right provides 'Debate Info' including the total score (3), arguments (3), and total votes (3). At the bottom, there's an advertisement for 'MAGNISES' and social media sharing options.

The screenshot shows a web browser window for the Wikipedia page of Tony Blair. The page includes a large portrait of him, his biography, and a sidebar with sections like 'Prime Minister of the United Kingdom' and 'Leader of the Opposition'. The main text discusses his political career, including his role as Prime Minister from 1997 to 2007 and his leadership of the Labour Party. It also mentions his policies and controversies, such as the introduction of the National Minimum Wage Act and the Good Friday Agreement.

Re: essence documentation - Message (Plain Text)

File Edit View Insert Format Tools Actions Help

Reply All Forward A⁺ B⁻ C⁰ D¹ E² F³ G⁴ H⁵ I⁶ J⁷ K⁸ L⁹ X⁰ Y¹ Z²

Extra line breaks in this message were removed.

From: radev@si.umich.edu on behalf of Ali Hakim [shakim@mac.com]
To: sam
Cc: radev@umich.edu; Rohit Laugani
Subject: Re: essence documentation

Sent: Wed 8/10/2005 1:29 PM

OK. I have added Essence.tex to the doc directory in the CLAIR project.
Can someone tell me how to set up my login shell so I don't have to export the CVS root every time or use the full path to CVS? Thanks.

Ali

On Aug 10, 2005, at 3:05 PM, sam wrote:

```
>
> In message <20050810185818.D11E0B84918tangra.si.umich.edu>,
> radev@umich.edu writes:
>
>> sam wrote:
>>>
>>> I will try to make a docs directory in the CVS repository.
>>> (I did this once. I think I just used cvs add with a directory just
>>> as I would have for a file.... )
>>>
>>> Should it be parallel to the CLAIR directory or below it?
>>>
>>>
>>> Below it.
>>>
>>>
>> OK, it's there now.
>
```

CNN.com - Turmoil in Peru as PM quits - Aug 11, 2005 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.cnn.com/2005/WORLD/americas/08/11/peru.toledo.reut/index.html

Fox News CNN.com Member Services

SEARCH THE WEB CNN.com SEARCH Powered by YAHOO! search

WORLD

Turmoil in Peru as PM quits

Thursday, August 11, 2005; Posted: 8:01 p.m. EDT (00:01 GMT)

LIMA, Peru (Reuters) -- Peruvian President Alejandro Toledo said Thursday he had asked all his ministers to resign, their resignations and would evaluate who would stay on in their jobs.

The government was plunged into crisis earlier when Prime Minister Carlos Ferrero quit following the appointment of Toledo's controversial top ally, Fernando Olvera, as foreign minister.

Housing and Construction Minister Carlos Bruce also resigned immediately.

Bruce and Ferrero both publicly split with Olvera over legalizing some cultivation of the raw material for cocaine, and their resignations signal more turbulence ahead for the unpopular Toledo in his final months in office.

Under Peru's constitution, once a prime minister resigns all ministers must tender their resignations.

"Carlos Ferrero has tendered his irrevocable resignation as prime minister," a statement from Toledo's office said.

Ferrero, a veteran lawmaker who has been prime minister since December 2003, would have had to quit the Cabinet by early October if he wanted to run for a congressional seat, but the timing of his announcement may have been after Toledo's closest ally was shown in as foreign minister -- was a surprise.

President Alejandro Toledo has a low public approval rating.

Want to know if a thief is using your credit card?

Advertiser links: [Compare August Mortgage Rates](#) [Get a \\$50,000 loan for \\$72 per month.](#) [Restaurant where rates are low.](#) [www.lowermybills.com](#)

MyCashNow - \$100 - \$1,500 Overnight

Payday Loan Cash goes in your account overnight. Very low fees. Fast decisions... [www.mycashnow.com](#)

Comcast High-Speed Internet

Order today for a \$19.99/mo. special, free modem, plus get \$75 cash back when... [www.comcastoffers.com](#)

Florist.com Save 10% - Fresh Flowers

Crooked Timber Mozilla Thunderbird

File Edit View Go Message Tools Help

Get Mail Write Address Book Reply Reply All Forward Delete Junk Print Stop

Folders

News & Blogs

- Trash
- Debtors' barb is overrated
- Crooked Timber
- Ernie's 3D Pancakes
- Geeking with Greg (3)
- Hit and Run (160)
- Interesting People (200)
- Salon.com (324)
- State Magazine (402)
- The Long Tail (17)
- The Technor...eblog (17)
- Local Folders
- Unsent Messages
- Drafts
- Sent
- Trash

View: All

Subject: Spreading Statistics, cont.

From: Ted

Date: 8/4/2005 9:06 AM

Website: <http://crookedtimber.org/2005/08/04/spreading-statistics-cont/>

I noted a few days ago that Senator Rick Santorum made a claim in an online interview about federal taxation. Senator Santorum said that the federal tax rate for the average family has gone up from 2% (in 1950) to 27% today. Furthermore, he claimed that income from a second worker simply replaces the money that the family pays in increased federal taxes. They would enjoy the same net income if taxes went back to 1950 levels and the second worker stayed at home.

I'm really rather sure that this isn't true. I'm relying on the [Tax Policy Center](#). They say that federal taxes on a family of four at the median income have gone up from about 7.4% to about 14.4%, and that the family would have saved \$4436 if we could roll back tax rates. That doesn't correspond to the Senator's story.

I checked last night, and Santorum repeats this point in his book, *It Takes a Family*. It's on page 123 and 124, and there's no source. (There's a bibliography of sorts, but it just lists a series of sources used in each section. There's no way to connect any specific point to any source.) When I called his press office again to ask for a source, they referred me to the publisher, who couldn't help me. Nonetheless, he's repeated this claim at least two more times, on [Hardball with Chris Matthews](#) and on [Fox News](#).

Shouldn't the Senator care whether what he's saying is right or wrong? Wouldn't it be nice if a journalist asked him about it?

Incidentally, is there anything more depressing than the "Current Events" section of a modern-day bookstore? There are so many rows of hastily-written, 200-250 page books with giant print, huge margins, and a cover featuring a smug bastard under a title like "THEY'RE ALL AGAINST YOU: How Hollywood, the French, and the CIA Have Conspired to Pollute Your Precious Bodily Fluids and What You Can Do To Stop Them." Robert Bork's *Slouching Towards Gomorrah* looks like Winston Churchill in all that dreck.)

Unread: 0 Total: 118

Dell 2Q Profit Rises, Revenue Misses View - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://biz.yahoo.com/ep/050811/earns_dell.html?v=7

YAHOO! FINANCE

Sign In New User? Sign Up

Finance Home - Help AP Associated Press

Welcome [Sign In]

Financial News

Enter symbol(s) Basic Get Symbol Lookup

Scottrade More Speed. More Power. Still Just \$7. **HARRISdirect \$100 CREDIT** **GET 100 COMMISSION FREE TRADES ON TRADE SECURITIES**

Related Quote

Dell 2Q Profit Rises, Revenue Misses View

Thursday August 11, 7:01 pm ET By Matt Slagle, AP Technology Writer

Dell Second-Quarter Profit Rises 28 Percent, but Revenue Misses Wall Street Expectations

DALLAS (AP) -- Increased shipments boosted Dell Inc.'s second-quarter by 28 percent but revenue fell below Wall Street expectations as the company blamed overly aggressive pricing of its low-end desktop and notebook computers.

Revenue at the world's largest PC maker rose 15 percent to \$13.43 billion from \$11.71 billion last year.

The results reported Thursday matched analyst forecasts of 38 cents per share but fell below projected revenue of \$13.71 billion, according to Thomson Financial.

Dell shares, which fell 15 cents to close at \$39.58 on the Nasdaq Stock Market, tumbled 7.7 percent, or \$3.03, to \$36.55 in late trading. Dell released the profit report after the market close.

Net income climbed to \$1.02 billion, or 41 cents per share, from \$799 million, or 31 cents per share, during the year-ago period.

The Round Rock, Texas-based company said its results included a 3-cent tax benefit, without which it would have earned 38 cents per share in the latest quarter.

Chief Executive Kevin Rollins blamed the revenue shortfall on lower prices on its low-end desktops and laptops.

Advertisement

3.50% Annual Percentage Yield

America's Highest Savings Rate Just Got Higher!

EmigrantDirect.com™ More money for your money™

Member FDIC Learn More

Top Stories

- Dow Rises 91, Nasdaq Up 17 on Late Rally - AP (5:44 pm)**
- Ex-WorldCom CFO Gets Five Years in Prison - AP (5:41 pm)**
- July Retail Sales Climb on Auto Purchases - AP (5:00 pm)**

Google Press Center: Press Release - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

<http://www.google.com/press/pressrel/univision.html>

Press Center

Univision.Com And Google Partner To Bring Search Services to Nation's Most Visited Spanish-Language Website

MOUNTAIN VIEW, Calif. and NEW YORK – July 28, 2005 – Univision Online, Inc., the interactive division of Univision Communications Inc. (NYSE: UVN), today announced that it has entered into multi-year partnership with Google Inc. (NASDAQ: GOOG). The partnership combines the power of the most visited Spanish-language website with an industry leader in search technology and advertising.

Through the partnership with Google, visitors to Univision.com will now receive more relevant, comprehensive, and timely Spanish-language search results optimized for quick delivery in a user-friendly environment. The relationship also connects Univision.com users with more products and services online through targeted advertisements on the search and content pages of the website. In addition, these ads increase the visibility of websites and advertisers serving Hispanics.

"As the most-visited Spanish-language website, Univision.com is excited to work with Google to bring the benefits of more targeted search capabilities to our users," said Javier Saralegui, President, Univision Online. "In addition, for those consumer companies not yet targeting Hispanics online, this partnership should provide additional impetus for them to develop websites in Spanish."

By combining the No.1 Spanish-language website with sophisticated and effective search technology, Univision.com and Google are working together to lead in ensuring that search becomes a more prominent functionality and effective targeting method in the Hispanic online marketplace.

"Google's agreement with the largest Spanish-language media company in the United States provides Univision's users with relevant information from search results and ads and gives our advertisers a new and targeted way to reach the Hispanic audience," said Tim Armstrong, Vice President, Advertising Sales, Google Inc.

About Univision Online

Univision Online, Inc. is the interactive division of Univision Communications Inc. (NYSE: UVN), the premier Spanish-language media company in the United States. Univision Online is the leading provider of the primary Spanish-language television network in the U.S. located at www.univision.com, and is a wholly owned subsidiary of Univision Communications Inc. Its operations include Univision Network, the most-watched Spanish-language broadcast television network in the U.S. reaching 96% of U.S. Hispanic Households; TeleFutura Network, a general-interest Spanish-language broadcast television network, which was launched in 2002 and now reaches 85% of U.S. Hispanic Households; Univision Television Group, which owns and operates 27 television network television stations and 1 non-network television stations; Univision Radio, which owns and operates 110 television stations and 10 radio stations; Univision Galvisión, the country's leading Spanish-language cable network; Univision Radio, the leading Spanish-language radio network, which owns and/or operates 66 radio stations in 16 of the top 25 U.S. Hispanic markets; and 4 stations in Puerto Rico; Univision Music Group, which includes Univision Records, Fonovisa Records, and a 50% interest in Mexico-based Discs Records labels as well as Fonomusic and America Music Publishing companies. Univision Communications also has a 50% interest in TuTV, a joint venture formed to broadcast Televisa's pay television channels in the U.S., and a non-voting 27% interest in Entravision Communications Corporation, a public Spanish-language media company. Univision Communications is headquartered in Los Angeles with television network operations in Miami and television and radio stations and sales offices in major cities throughout the United States.

For more information, please visit www.univision.net.

About Google Inc.

Google's innovative search technologies connect millions of people around the world with information every day. Founded in 1995 by Stanford Ph.D. students Larry Page and Sergey Brin, Google today is a top web property in all major global markets. Google's targeted advertising program provides businesses of all sizes with measurable results while enhancing the overall web experience for users.

Done

2004_ibm_annual.pdf (application/pdf Object) - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://ibm8k.filing.internationalservicesmachines.com/2004_ibm_annual.pdf

INTERNATIONAL SERVICES MACHINES

IBM ANNUAL REPORT 2004

1 of 100

PTEK HOLDINGS INC (Form: 8-K, Received: 09/07/2004 17:03:57) - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

<http://www.shareholder.com/common/edgar/680604/1193125-04-152871/04-00.html>

**UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, DC 20549**

FORM 8-K

**CURRENT REPORT PURSUANT
TO SECTION 13 OR 15(D) OF THE
SECURITIES EXCHANGE ACT OF 1934**

Date of report (Date of earliest event reported) September 7, 2004

PTEK HOLDINGS, INC.
(Exact Name of Registrant as Specified in Its Charter)

GEORGIA
(State or Other Jurisdiction of Incorporation)

000-27778
(Commission File Number)

50-3074176
(IRS Employer Identification No.)

3399 Peachtree Road, NE, Suite 700, Atlanta, Georgia 30326
(Address of Principal Executive Offices) (Zip Code)

404-262-8400
(Registrant's Telephone Number, Including Area Code)

Check the appropriate box below if the Form 8-K filing is intended to simultaneously satisfy the filing obligation of the registrant under any of the following provisions (see General Instruction A.2. below):

- Written communications pursuant to Rule 425 under the Securities Act (17 CFR 230.425)
- Soliciting material pursuant to Rule 14a-12 under the Exchange Act (17 CFR 240.14a-12)
- Pre-commencement communications pursuant to Rule 14d-2(b) under the Exchange Act (17 CFR 240.14d-2(b))
- Pre-commencement communications pursuant to Rule 13e-4(c) under the Exchange Act (17 CFR 240.13e-4(c))

Check the appropriate box below if the Form 8-K filing is intended to simultaneously satisfy the filing obligation of the registrant under any of the following provisions (see General Instruction A.2. below):

- Written communications pursuant to Rule 425 under the Securities Act (17 CFR 230.425)
- Soliciting material pursuant to Rule 14a-12 under the Exchange Act (17 CFR 240.14a-12)
- Pre-commencement communications pursuant to Rule 14d-2(b) under the Exchange Act (17 CFR 240.14d-2(b))
- Pre-commencement communications pursuant to Rule 13e-4(c) under the Exchange Act (17 CFR 240.13e-4(c))

Done

Tony Blair - Wikipedia, the free encyclopedia - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://en.wikipedia.org/wiki/Tony_Blair

WIKIPEDIA
The Free Encyclopedia

article discussion edit this page history

Wikimedia needs your help in its 21-day fund drive. See our fundraising page.
Over US\$185,000 has been donated since the drive began on 19 August. Thank you for your generosity!

Tony Blair

From Wikipedia, the free encyclopedia.

The Right Honourable Anthony Charles Lynton Blair (born May 6, 1953 in Edinburgh, Scotland) is the current Prime Minister of the United Kingdom. He has led the Labour Party since July 1994, following the death of John Smith in May of that year and brought Labour into power with a landslide victory in the 1997 general election, replacing John Major as Prime Minister and ending 18 years of Conservative government. He is now the Labour Party's longest-serving Prime Minister, and the only person to have led the party to three consecutive general election victories, just as Margaret Thatcher was the only Conservative Prime Minister to win three consecutive general elections.

Blair moved the Labour Party towards the centre of British politics, using the term "New Labour" to distinguish what he identifies as "modern social democracy" and his party's refusal to reverse privatisation and support for a market economy from its past belief in nationalisation and Fabian socialism. However, critics on the left feel that he has compromised the principles of the founders of the Labour party, and that the Blair government has moved too far to the right, placing insufficient emphasis on traditional Labour priorities such as the redistribution of wealth.

Contents

- 1 Early and private life
- 2 Begins political career
- 3 In opposition
- 4 First term 1997 to 2001
- 5 Second term 2001 to 2005
- 6 Third term 2005 to present
- 7 Caricature and satire of Blair
- 8 See also
- 9 References
- 10 Further reading

The Rt Hon. Tony Blair

Appointed PM 2 May 1997

PM predecessor John Major

Date of birth 6 May 1953

Place of birth Edinburgh, Scotland

Political Party Labour

Constituency Sedgefield

Since the advent of the War on Terror, much of the Prime Minister's political agenda has been dominated by foreign affairs, particularly those concerning Iraq, and he has supported many aspects of the foreign policy of United States president George W. Bush, sending British forces to participate in the 2003 invasion of Iraq and the subsequent occupation and conflict. Blair's Labour party still managed to win an unprecedented third term in the 2005 general election (unprecedented for the Labour Party; the Conservatives having won three consecutive terms twice since the Second World War). Although Labour's majority in the House of Commons was reduced considerably to 67 MPs, this remains a substantial working majority and a measure of the relative political weakness and poor credibility of the main opposition party, the Conservatives. Some journalists immediately noticed that New Labour no longer has a majority independently of Old Labour, although this may discourage some MPs from mounting damaging rebellions against the leadership.

While Blair is in no danger of losing a potential vote of no confidence, the fall in the Labour vote (from 41% to 35%) has renewed speculation amongst commentators as to how long his leadership can continue. It is widely predicted that he will be succeeded by his ambitious Chancellor of the Exchequer Gordon Brown before the next General Election (which will occur at the latest in 2010).

Early and private life

Blair was born in Edinburgh, Scotland. His father, Leo Blair, was a barrister and later a law lecturer who, having been a communist in his youth, was active in the Conservative Party. Leo Blair had ambitions to stand for Parliament in Durham, which were thwarted when he had a stroke when his son was 11, an event which affected

Done

Induction of Influenza-Specific Mucosal Immunity by an Attenuated Recombinant Sendai Virus

Thuc-vy L. Le¹, Elena Mironova², Dominique Garcin², Richard W. Compans^{1*}

1 Department of Microbiology and Immunology and Emory Vaccine Center, Emory University School of Medicine, Atlanta, Georgia, United States of America,

2 Department of Microbiology and Molecular Medicine, University of Geneva School of Medicine, Geneva, Switzerland

Recent advances in molecular genetics have permitted the development of novel virus-based vectors for the delivery of genes and expression of gene products [6,7,8]. These live vectors have the advantage of promoting robust immune responses due to their ability to replicate, and induce expression of genes at high efficiency. Sendai virus is a member of the *Paramyxoviridae* family, belongs in the genus respirovirus and shares 60–80% sequence homology to human parainfluenza virus type 1 (HPIV-1) [9,10]. The viral genome consists of a negative sense, non-segmented RNA. Although Sendai virus was originally isolated from humans during an outbreak of pneumonitis [11] subsequent human exposures to Sendai virus have not resulted in observed pathology [12]. The virus is commonly isolated from mouse colonies and Sendai virus infection in mice leads to bronchopneumonia, causing severe pathology and inflammation in the respiratory tract. The sequence homology and similarities in respiratory pathology have made Sendai virus a mouse model for HPIV-1. Immunization with Sendai virus promotes an immune response in non-human primates that is protective against HPIV-1 [13,14] and clinical trials are underway to determine the efficacy of this virus for protection against HPIV-1 in humans [15]. Sendai virus naturally infects the respiratory tract of mice and recombinant viruses have been reported to efficiently transduce luciferase, lac Z and green fluorescent protein (GFP) genes in the airways of mice or ferrets as well as primary human nasal epithelial cells [16]. These data support the hypothesis that intranasal (i.n.) immunization with a recombinant Sendai virus will mediate heterologous gene expression in mucosal tissues and induce antibodies that are specific to a recombinant protein. A major advantage of a recombinant Sendai virus based vaccine is the observation that recurrence of parainfluenza virus infections is common in humans [12,17] suggesting that anti-vector responses are limited, making repeated administration of such a vaccine possible.

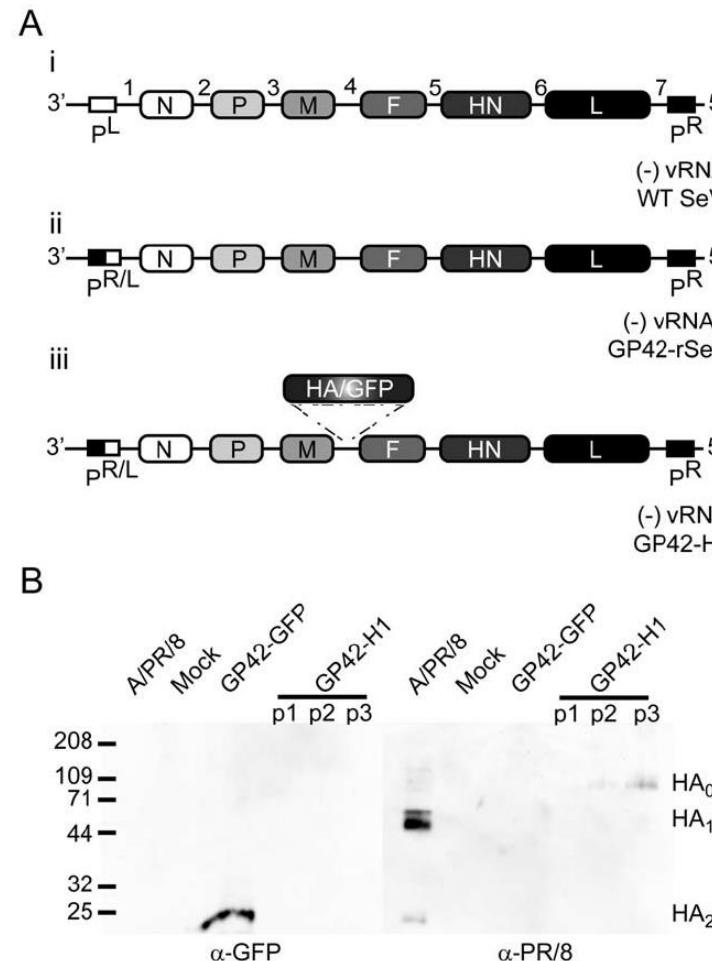


Figure 1. Generation of the GP42-H1 vector. A) Representation of wild-type Sendai virus gene construct (i) illustrating the major viral genes N, P, M, F, HN and L. The left promoter (P^L) and right promoter (P^R) function as the genomic promoter and anti-genomic promoter respectively. Seven gene boundaries that encode the conserved regulatory transcription start and transcription stop sequence are represented numerically. The mutant GP42-SeV (ii) genomic RNA is identical to WT Sendai virus with the exception of 3' P^L in which 42 nucleotides of the P^L were replaced with the corresponding sequence from P^R ($P^{R/L}$). Additional transcription start, stop, poly-adenylation sequences and a unique Mlu I restriction site were introduced into the intergenic region between the Sendai M and F genes. Using the unique Mlu I restriction site, the GFP or HA gene was inserted (respecting the rule of six) generating the recombinant Sendai GP42-GFP [19] or GP42-H1 vectors (iii). B) Recombinant GP42-H1 virus was cultured in BSR-T7 cells for three passages. Cell free supernatant containing virus suspensions were collected and used to infect CV-1 cells (refer to Materials and Methods). Proteins from GP42-GFP or GP42-H1 infected cell extracts were resolved on SDS-PAGE and screened for GFP (left) or HA (right) expression by western analysis. Mock infected cells and allantoic fluid from PR/8 infected eggs are also shown.
doi:10.1371/journal.pone.0018780.g001

Plos ONE

DOI: 10.1371/journal.pone.0018780

Recent advances in molecular genetics have permitted the development of novel virus-based vectors for the delivery of genes and expression of gene products [6,7,8]. These live vectors have the advantage of promoting robust immune responses due to their ability to replicate, and induce expression of genes at high efficiency. **Sendai virus** is a member of the Paramyxoviridae family, belongs in the genus respirovirus and shares 60–80% sequence homology to **human parainfluenza virus type 1 (HPIV-1)** [9,10].

The viral genome consists of a negative sense, non-segmented RNA. Although **Sendai virus** was originally isolated from humans during an outbreak of pneumonitis [11] subsequent human exposures to **Sendai virus** have not resulted in observed pathology [12]. The virus is commonly isolated from mouse colonies and Sendai virus infection in mice leads to bronchopneumonia, causing severe pathology and inflammation in the respiratory tract. The sequence homology and similarities in respiratory pathology have made Sendai virus a mouse model for HPIV-1. Immunization with Sendai virus promotes an immune response in non-human primates that is protective against **HPIV-1** [13,14] and clinical trials are underway to determine the efficacy of this virus for protection against HPIV-1 in humans [15]. Sendai virus naturally infects the respiratory tract of mice and **recombinant viruses have been reported to efficiently transduce luciferase, lac Z and green fluorescent protein (GFP) genes in the airways of mice or ferrets as well as primary human nasal epithelial cells** [16].

These data support the hypothesis that intranasal (i.n.) immunization with a recombinant Sendai virus will mediate heterologous gene expression in mucosal tissues and induce antibodies that are specific to a recombinant protein. A major advantage of a recombinant Sendai virus based vaccine is the observation that recurrence of **parainfluenza virus** infections is common in humans [12,17] **suggesting** that anti-vector responses are limited, making repeated administration of such a vaccine possible.

Named entities + variants (**human parainfluenza virus type, HPIV-1**)

Speculation (**reported, suggesting**)

Species (**human**)

Cell types (**nasal epithelial cells**)

Facts

References

Electronic Health Records

TITLE: PC ACUTE CARE VISIT	ENTRY DATE: FEB 04, 2000@11:20
DATE OF NOTE: FEB 04, 2000@11:18	EXP COSIGNER:
AUTHOR:	STATUS: COMPLETED
URGENCY:	

Chief Complaint: Patient notes 1 month history of blurred vision and frequent urination

HISTORY OF PRESENT ILLNESS:

DEMO,FATHER is a 44 year-old MALE who presents complaining of blurred vision for the past 1 month. He finds it is difficult for him to read clearly and is even effecting his driving. He also notes that he has been getting up to the bathroom frequently, esp. at night. He now routinely goes to urinate 3-4 times a night. He is not aware of any particular weight loss, but does feel thirsty much of the time.

PAST MEDICAL HISTORY:

Illnesses: Hypertension

Surgeries: None

Allergies: PENICILLINS

Medications:

1) HYDROCHLOROTHIAZIDE 25MG TAB** Qty: 45 ACTIVE
for 90 days Sig: TAKE ONE-HALF TABLET MOUTH EVERY MORNING THC BLOOD PRESSURE Refills: 0

2) METOPROLOL 25MG XL TAB Qty: 90 for 90 ACTIVE
days Sig: TAKE ONE TABLET MOUTH QDAY FOR THE HEART Refills: 0

FAMILY HISTORY:

Diabetes Father, Sibling, Grandparent

Literary Texts

- Project Gutenberg (<http://www.gutenberg.org/browse/scores/top>)
- A team of horses passed from Finglas with toiling plodding tread, dragging through the funereal silence a creaking waggon on which lay a granite block. The waggoner marching at their head saluted.
 - Ulysses - <http://www.gutenberg.org/files/4300/4300-h/4300-h.htm>
- There was no possibility of taking a walk that day. We had been wandering, indeed, in the leafless shrubbery an hour in the morning; but since dinner (Mrs. Reed, when there was no company, dined early) the cold winter wind had brought with it clouds so sombre, and a rain so penetrating, that further out-door exercise was now out of the question.
 - Jane Eyre - <http://www.gutenberg.org/files/1260/1260-h/1260-h.htm>
- Dorothy lived in the midst of the great Kansas prairies, with Uncle Henry, who was a farmer, and Aunt Em, who was the farmer's wife. Their house was small, for the lumber to build it had to be carried by wagon many miles. There were four walls, a floor and a roof, which made one room; and this room contained a rusty looking cookstove, a cupboard for the dishes, a table, three or four chairs, and the beds. Uncle Henry and Aunt Em had a big bed in one corner, and Dorothy a little bed in another corner. There was no garret at all, and no cellar--except a small hole dug in the ground, called a cyclone cellar, where the family could go in case one of those great whirlwinds arose, mighty enough to crush any building in its path. It was reached by a trap door in the middle of the floor, from which a ladder led down into the small, dark hole.
 - The Wizard of Oz - <http://www.gutenberg.org/files/55/55-h/55-h.htm>

A Really Long Literary Sentence

- Try parsing this
 - “Bloat is one of the co-tenants of the place, a maisonette erected last century, not far from the Chelsea Embankment, by Corydon Throsp, an acquaintance of the Rossettis' who wore hair smocks and liked to cultivate pharmaceutical plants up on the roof (a tradition young Osbie Feel has lately revived), a few of them hardy enough to survive fogs and frosts, but most returning, as fragments of peculiar alkaloids, to rooftop earth, along with manure from a trio of prize Wessex Saddleback sows quartered there by Throsp's successor, and dead leaves off many decorative trees transplanted to the roof by later tenants, and the odd unstomachable meal thrown or vomited there by this or that sensitive epicurean-all got scumbled together, eventually, by the knives of the seasons, to an impasto, feet thick, of unbelievable black topsoil in which anything could grow, not the least being bananas.”
- Do you know the source?

Quiz Answer

- “Gravity’s Rainbow” (by Thomas Pynchon), known for its use of very arcane words and complicated sentence (and plot) structure.
- Another such work is “Finnegans Wake” by James Joyce.
- Poetry is even more difficult.

Natural Language Processing

Research Projects and Evaluations

Demos

- <https://playground.tensorflow.org/>
- <https://p.migdal.pl/interactive-machine-learning-list/>
- <https://demo.allennlp.org/reading-comprehension>
- <https://allenai.org/demos/>
- <http://aristo-demo.allenai.org/>
- <http://text-processing.com/demo/>
- <https://explosion.ai/demos/>
- <https://huggingface.co/hmtl/>
- <https://corenlp.run/>
- <http://nlp.stanford.edu:8080/corenlp/>
- <https://talktotransformer.com/>
- <https://cs.stanford.edu/people/karpathy/convnetjs/>

State of the Art NLP Challenges

<https://gluebenchmark.com/leaderboard>

<https://leaderboard.allenai.org/arc/submissions/public>

<https://yale-lily.github.io/spider>

<https://stanfordnlp.github.io/coqa/>

<https://rajpurkar.github.io/SQuAD-explorer/>

<http://www.msmarco.org/leaders.aspx>

<https://github.com/salesforce/decaNLP>

<http://ruder.io/tracking-progress-nlp/>

http://nlpprogress.com/english/question_answerering.html

<https://quac.ai/>

<https://www.tau-nlp.org/csqa-leaderboard>

<https://nlp.stanford.edu/software/sempre/>

<http://nlp.cs.washington.edu/triviaqa/>

<http://lic.nlp.cornell.edu/nlvr/>

<https://decanlp.com/>

<https://worksheets.codalab.org/worksheets/0xba659fe363cb46e7a505c5b6a774dc8a/>

<http://fever.ai/task.html>

<http://www.fakenewschallenge.org/>

<https://www.nyu.edu/projects/bowman/multinli/>

<http://convai.io/>

<http://sharedtask.duolingo.com/>

YOU ASKED ARISTO:

Fossil fuels were formed from

- (A) volcanoes
- (B) the remains of living things
- (C) gases in the atmosphere
- (D) water trapped inside rocks

ARISTO ANSWERED:

Question: Fossil fuels were formed from

Aristo's Answer: (B) the remains of living things

Correct Answer: B

Confidence: 84.13%

as computed from these reasoners:

Information Retrieval: 50.50% [MORE INFO](#)

Justification Sentence: Fossil fuels were formed from the fossilized remains of dead plants and animals that once lived millions of years ago.

Table Reasoning: 34.70% [MORE INFO](#)

Knowledge Used: [living organisms | some form of respiration]

Topic Matching: 19.37% [MORE INFO](#)

Topic: fossil

Tuple Reasoning: 45.87% [MORE INFO](#)

Knowledge Used: [Fuels | come | from the remains of living things] [Fossil fuel | is formed | by the remains of plants and animals] [Fossil fuels | are formed | from the remains of once living organisms]

AristoRoBERTa: 86.98% [MORE INFO](#)

Ethnicity	Islam	Christianity	Judaism	Buddhism	Other	None / Atheism	n/a
TOTALS	70.20%	26.32%	0.03%	0.09%	0.02%	2.82%	0.51%
Kazakh	98.34%	0.39%	0.02%	0.01%	0.02%	0.98%	0.26%
Russian	1.43%	91.64%	0.04%	0.02%	0.03%	6.09%	0.75%
Uzbek	99.05%	0.39%	0.01%	0.01%	0.02%	0.37%	0.16%
Ukrainian	0.94%	90.74%	0.03%	0.01%	0.02%	7.31%	0.94%
Uyghur	98.35%	0.51%	0.02%	0.01%	0.03%	0.61%	0.47%
Tatar	79.57%	10.24%	0.02%	0.03%	0.06%	8.11%	1.97%
German	1.58%	81.59%	0.05%	0.04%	0.11%	13.96%	2.68%
Korean	5.24%	49.35%	0.21%	11.40%	0.14%	28.51%	5.16%
Turkish	99.13%	0.30%	0.01%	0.01%	0.02%	0.33%	0.21%
Azeri	94.81%	2.51%	0.02%	0.02%	0.03%	1.86%	0.76%
Belorussian	0.79%	90.16%	0.04%	0.01%	0.03%	7.82%	1.15%
Dungan	98.93%	0.37%	0.01%	0.03%	0.04%	0.34%	0.28%
Kurdish	98.28%	0.53%	0.03%	0.02%	0.02%	0.74%	0.38%
Tajik	97.78%	0.91%	0.01%	0.02%	0.08%	0.85%	0.35%
Polish	0.69%	90.07%	0.04%	0.01%	0.13%	7.30%	1.76%
Chechen	93.69%	2.99%	0.02%	0.01%	0.05%	2.08%	1.16%
Kyrgyz	96.67%	0.89%	0.03%	0.03%	0.02%	1.51%	0.86%
Others	34.69%	52.32%	0.82%	0.91%	0.13%	8.44%	2.69%

URL <http://en.wikipedia.org/wiki?action=render&curid=16642&oldid=602424190>

Title Kazakhstan

Table # 0

nt-83

which ethnicity is previous from dungan

Belorussian

nt-1833

which ethnicity has more followers of islam: tatar or tajik?

Tajik

nt-2689

are there more christian russians or ukrainians in kazakhstan?

Russian

nt-3253

how many times is each religion listed?

once

nt-4590

which ethnicity is above german

Tatar

nt-5037

what is the difference in percentage between korean buddists and german buddists?

11.36%

nt-6191

which ethnicity has the most buddhists in kazakhstan?

Korean

nt-6739

which ethnicity is first on the chart

Kazakh

nt-10453

what ethnicity is next under belorussian on the chart?

Dungan



Spider 1.0



Yale Semantic Parsing and Text-to-SQL Challenge

What is Spider?

Spider is a large-scale *complex and cross-domain* semantic parsing and text-to-SQL dataset annotated by 11 Yale students. The goal of the Spider challenge is to develop natural language interfaces to cross-domain databases. It consists of 10,181 questions and 5,693 unique complex SQL queries on 200 databases with multiple tables covering 138 different domains. In Spider 1.0, different complex SQL queries and databases appear in train and test sets. To do well on it, systems must *generalize well to not only new SQL queries but also new database schemas*.

Why we call it "Spider"? It is because our dataset is complex and cross-domain like a spider crawling across multiple complex(with many foreign keys) nests(databases).

[Spider Paper \(EMNLP'18\)](#)
[Spider Post](#)

Related challenges: multi-turn [SParC](#) and conversational [CoSQL](#) text-to-SQL tasks.

[SParC Challenge \(ACL'19\)](#)
[CoSQL Challenge \(EMNLP'19\)](#)

Leaderboard - Exact Set Match without Values

For exact matching evaluation, instead of simply conducting string comparison between the predicted and gold SQL queries, we decompose each SQL into several clauses, and conduct set comparison in each SQL clause. Please refer to the paper and [the Github page](#) for more details.

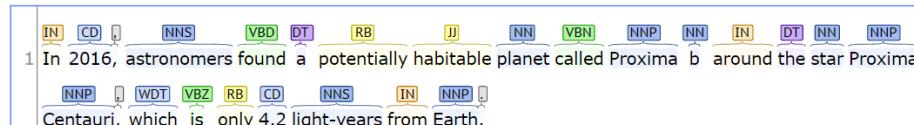
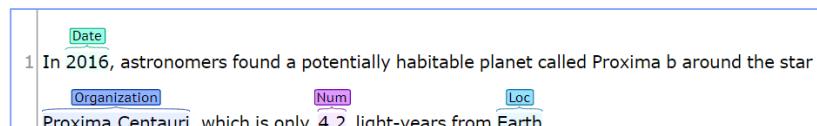
Rank	Model	Dev	Test
1	RATSQL v2 + BERT (DB content used) <i>Anonymous</i>	65.8	61.9
Dec 13, 2019			
2	IRNet++ + XLNet (DB content used) <i>Anonymous</i>	65.5	60.1
Dec 18, 2019			
3	RYANSQ + BERT <i>Anonymous</i>	66.6	58.2
Nov 12, 2019			
4	RATSQL v2 (DB content used) <i>Anonymous</i>	62.7	57.2
Dec 13, 2019			
5	RASQL + BERT <i>Anonymous</i>	60.8	55.7
Dec 13, 2019			
6	EditSQL+LSL + BERT <i>Anonymous</i>	57.9	55.2
Dec 13, 2019			
7	IRNet v2 + BERT <i>Microsoft Research Asia</i>	63.9	55.0
June 24, 2019			
8	GIRN + BERT <i>Anonymous</i>	60.2	54.8
Sep 20, 2019			

Stanford CoreNLP

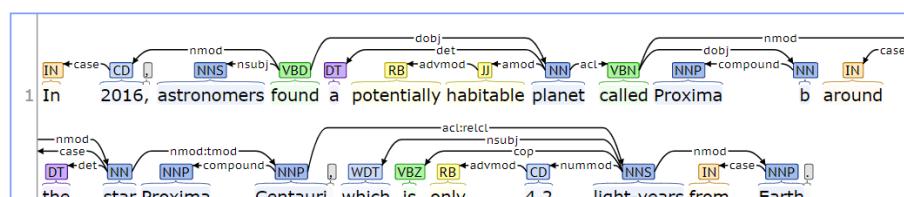
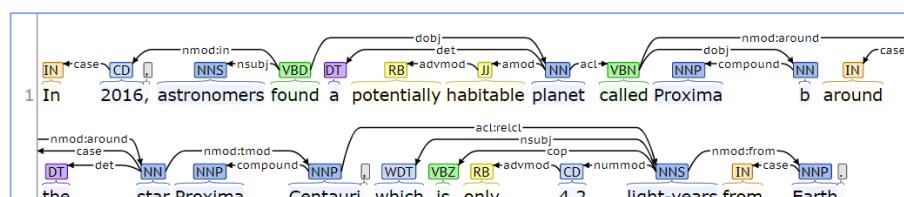
Output format: Visualise ▾

Please enter your text here:

In 2016, astronomers found a potentially habitable planet called Proxima b around the star Proxima Centauri, which is only 4.2 light-years from Earth.

 Part-of-Speech:**Named Entity Recognition:****Coreference:**

1 In 2016, astronomers found a potentially habitable planet called Proxima b around the star Proxima Centauri, which is only 4.2 light-years from Earth.

Basic Dependencies:**Enhanced Dependencies:**

The LILY Lab

- <http://lily.yale.edu>
- <http://aan.how>
- <http://www.nacloweb.org>

LILY Projects

- Text summarization
 - Crosslingual information retrieval
 - Analysis of electronic health records
 - Survey generation
 - Text to SQL
 - Dialogue systems
 - Multilingual computing
 - Text generation
-
- **Ask me for details**

NLP research in LILY

- **Multilingual information retrieval** -- We collaborate with researchers from Columbia University, the University of Maryland, the University of Edinburgh, and the University of Cambridge to build search engines for English users to query documents written in other languages. This cross-lingual information retrieval system improves our capability of understanding and processing different low-resource languages and it offers users a reliable access to foreign documents. Currently working on Tagalog, Swahili, Somali, Lithuanian, Bulgarian, Pashto.
- **Resources for learning NLP and AI** -- We aim to make dynamic research topics more accessible to the public by generating surveys of topics, discovering prerequisite relations among topics and recommending appropriate resources based on a given individual's education background and needs. We host a search engine, AAN (All About NLP) and tool which is available at <http://aan.how>.
- **Medical NLP** – We work on NLP on electronic health records. The tasks include abbreviation disambiguation, patient digital pairs, patient history record summarization, medical text segmentation and translation,. We are interested in how to transfer general knowledge to the medical domain and how to take good advantage of the limited high-quality data. We are also developing an NLP toolkit for the medical domain to perform tasks like named entity recognition and relation extraction.
- **Natural language database interfaces and dialogue systems** -- The goal of this project is to allow users with no SQL background to talk to relational databases directly using human language. This way, anyone can easily query and analyze a vast amount of data. We also aim to build conversational interfaces for even more natural information access, where the users participate in a conversation and the system takes the responsibility of choosing data sources and developing queries.
- **Text summarization and text generation** – single and multi-document, monolingual and multilingual, extractive and abstractive summarization, opinion summarization, etc.

Natural Language Processing

113

NACLO

NACLO and IOL

- The North American Computational Linguistics Open Competition (formerly Olympiad)
 - Competition held since 2007 in the USA and Canada
 - <http://www.nacloweb.org>
- Best individual US performers so far:
 - Adam Hesterberg (2007)
 - Rebecca Jacobs (2007-2009) – 3 team golds + 2 individual medals
 - Ben Sklaroff (2010)
 - Morris Alper (2011)
 - Alex Wade (2012, 2013) – 2 team golds + 2 individual golds + 1 individual silver
 - Tom McCoy (2013) – Yale grad
 - Darryl Wu (2012, 2014)
 - James Wedgwood (2015, 2016) – Yale senior
 - Brian Xiao (2017), Swapnil Garg (2018)
- Other strong countries:
 - Russia, UK, Netherlands, Poland, Bulgaria, South Korea, Canada, China, Sweden, Slovenia, Brazil, Taiwan
- IOL – the International contest
 - Since 2003
 - 2014 in China, 2015 in Bulgaria, 2016 in India, 2017 in Ireland, 2018 in Czechia, 2019 in Korea, 2021 in Latvia
 - <http://www.ioling.org>

Consider these phrases in Ancient Greek (in a Roman-based transcription) and their unordered English translations:

- (A) *ho tōn hyiōn dulos*
- (B) *hoi tōn dulōn cyrioi*
- (C) *hoi tu emporu adelphoi*
- (D) *hoi tōn onōn emporoi*
- (E) *ho tu cyriu onos*
- (F) *ho tu oicu cyrios*
- (G) *ho tōn adelphōn oicos*
- (H) *hoi tōn cyriōn hyioi*

- (1) the donkey of the master
- (2) the brothers of the merchant
- (3) the merchants of the donkeys
- (4) the sons of the masters
- (5) the slave of the sons
- (6) the masters of the slaves
- (7) the house of the brothers
- (8) the master of the house

C1. Place the number of the correct English translation in the space following each Greek sentence. Explain your answers!

C2. Translate into Ancient Greek:

the houses of the merchants;
the donkeys of the slave

Explain your answers!

To the right is a Japanese word written in the *tenji* ("dot characters") writing system. The large dots represent the raised bumps; the tiny dots represent empty positions.

karaoke ::::: :::::

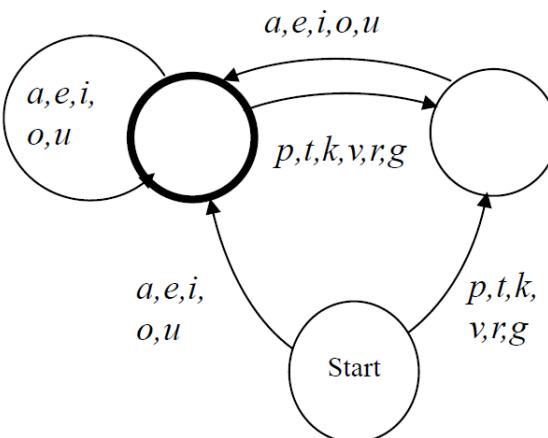
A1. The following *tenji* words represent *atari*, *haiku*, *katana*, *kimono*, *koi*, and *sake*. Which is which? You don't need to know either Japanese or Braille to figure it out; you'll find that the system is highly logical.

a. _____	::::: ::::	b. _____	::::: :::
c. _____	::::: ::::	d. _____	::::: ::::
e. _____	::::: :::	f. _____	::::: ::::

A2. What are the following words?

g. _____	::::: ::::	h. _____	::::: ::::
----------	------------	----------	------------

Here is a finite state automaton that can distinguish between possible and impossible words in Rotokas, a language spoken on the island of Bougainville off the coast of New Guinea. Rotokas has a very simple system of sounds and allows us to create a very small FSA.



An FSA works like a board game. Choose a word, and place your pencil on the space marked “Start”. Going through the letters of the word one at a time, move your pencil along the path marked with that letter. If the word ends and you’re at a space marked with a thicker circle, the word succeeds: it’s a possible Rotokas word! If the word ends and you’re not at a thicker circle, or you’re midway through the word and there’s no path corresponding to the next letter, the word fails: it’s *not* a possible Rotokas word!

Try it out with these possible and impossible words; the automaton should accept all the possible words and reject the impossible ones.

Possible Rotokas words	Impossible Rotokas words
tauo	grio
kareveiepa	ouag
puraveva	ovgi
ovokirovua	vonoka
avaopa	gataap
ouragaveva	oappa

II. Now, using the automaton above, put a check mark next to each possible Rotokas word:

- | | | |
|--------------------------------------|--|---|
| <input type="checkbox"/> <i>iu</i> | <input type="checkbox"/> <i>uente</i> | <input type="checkbox"/> <i>voav</i> |
| <input type="checkbox"/> <i>idau</i> | <input type="checkbox"/> <i>urioo</i> | <input type="checkbox"/> <i>uaia</i> |
| <input type="checkbox"/> <i>oire</i> | <input type="checkbox"/> <i>raorao</i> | <input type="checkbox"/> <i>oratreopaveiepa</i> |

On her visit to Armenia, Millie has gotten lost in Yerevan, the nation's capital. She is now at the Metropoliten (subway) station named Shengavit, but her friends are waiting for her at the station named Barekamutyun. Can you help Millie meet up with her friends?

1. Assuming Millie takes a train in the right direction, which will be the first stop after Shengavit?

Note that all names of stations listed below appear on the map.

- a. Gortsaranayin
- b. Zoravar Andranik
- c. Charbakh
- d. Garegin Njdehi Hraparak
- e. none of the above

2. After boarding at Shengavit, how many stops will it take Millie to get to Barekamutyun (don't include Shengavit itself in the number of stops)?



Some computational problems

<http://www.nacloweb.org/resources/problems/2016/N2016-B.pdf>
<http://www.nacloweb.org/resources/problems/2016/N2016-H.pdf>
<http://www.nacloweb.org/resources/problems/2016/N2016-K.pdf>
<http://www.nacloweb.org/resources/problems/2016/N2016-P.pdf>
<http://www.nacloweb.org/resources/problems/2015/N2015-E.pdf>
<http://www.nacloweb.org/resources/problems/2015/N2015-K.pdf>
<http://www.nacloweb.org/resources/problems/2015/N2015-M.pdf>
<http://www.nacloweb.org/resources/problems/2015/N2015-P.pdf>
<http://www.nacloweb.org/resources/problems/2015/N2015-G.pdf>
<http://www.nacloweb.org/resources/problems/2014/N2014-O.pdf>
<http://www.nacloweb.org/resources/problems/2014/N2014-P.pdf>
<http://www.nacloweb.org/resources/problems/2014/N2014-C.pdf>
<http://www.nacloweb.org/resources/problems/2014/N2014-J.pdf>
<http://www.nacloweb.org/resources/problems/2014/N2014-H.pdf>
<http://www.nacloweb.org/resources/problems/2014/N2014-L.pdf>
<http://www.nacloweb.org/resources/problems/2013/N2013-C.pdf>
<http://www.nacloweb.org/resources/problems/2013/N2013-F.pdf>
<http://www.nacloweb.org/resources/problems/2013/N2013-H.pdf>
<http://www.nacloweb.org/resources/problems/2013/N2013-L.pdf>
<http://www.nacloweb.org/resources/problems/2012/N2012-C.pdf>

<http://www.nacloweb.org/resources/problems/2013/N2013-N.pdf>
<http://www.nacloweb.org/resources/problems/2013/N2013-Q.pdf>
<http://www.nacloweb.org/resources/problems/2012/N2012-K.pdf>
<http://www.nacloweb.org/resources/problems/2012/N2012-O.pdf>
<http://www.nacloweb.org/resources/problems/2012/N2012-R.pdf>
<http://www.nacloweb.org/resources/problems/2011/F.pdf>
<http://www.nacloweb.org/resources/problems/2011/M.pdf>
<http://www.nacloweb.org/resources/problems/2010/D.pdf>
<http://www.nacloweb.org/resources/problems/2010/E.pdf>
<http://www.nacloweb.org/resources/problems/2010/I.pdf>
<http://www.nacloweb.org/resources/problems/2010/K.pdf>
<http://www.nacloweb.org/resources/problems/2009/N2009-E.pdf>
<http://www.nacloweb.org/resources/problems/2009/N2009-G.pdf>
<http://www.nacloweb.org/resources/problems/2009/N2009-J.pdf>
<http://www.nacloweb.org/resources/problems/2009/N2009-M.pdf>
<http://www.nacloweb.org/resources/problems/2008/N2008-F.pdf>
<http://www.nacloweb.org/resources/problems/2008/N2008-H.pdf>
<http://www.nacloweb.org/resources/problems/2008/N2008-I.pdf>
<http://www.nacloweb.org/resources/problems/2008/N2008-L.pdf>
<http://www.nacloweb.org/resources/problems/2007/N2007-A.pdf>
<http://www.nacloweb.org/resources/problems/2007/N2007-H.pdf>