DEEP LEARNING
IN NATURAL LANGUAGE PROCESSING

Paul COURSAUX
Lead AI Researcher - RIMINDER

DEEP LEARNING PRACTICAL COURSE
ECOLE POLYTECHNIQUE, 19/04/2018

Program & Course Logistics

- Course 1 : (05-04-18)
  - Introduction to Deep Learning - Mouhidine SEIV (Riminder)
- Course 2 : (12-04-18)
  - Deep Learning in Computer Vision - Slim FRIKHA (Riminder)
- **Course 3 : (19-04-18)**
  - **Deep Learning in NLP - Paul COURSAUX (Riminder)**
- Course 4 : (26-04-18)
  - Efficient Methods and Compression for Deep Learning - INVITED GUEST
- Course 5: (03-05-18)
  - Introduction to Deep Learning Frameworks - INVITED GUEST
- Course 6: (10-05-18)
  - Deployment in Production and Parallel Computing - INVITED GUEST

Location: Ecole Polytechnique from 6:30 pm to 7:30pm

riminder

https://github.com/riminder

Talk outline

I. *NLP overview*

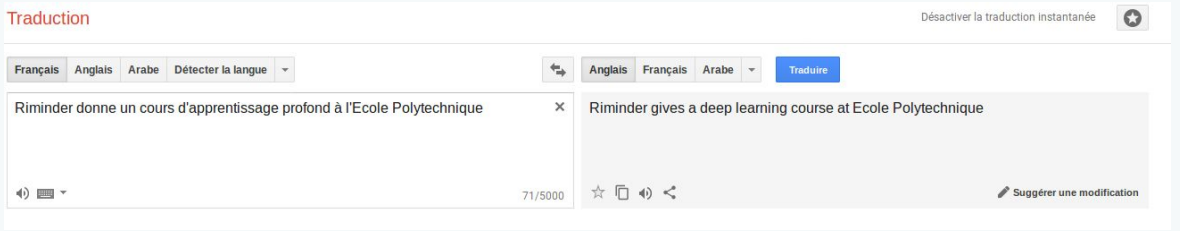
II. *From frequency to meaning: word vectors*

III. *Deep learning for text*

IV. *Text generation*

Why is NLP important ?

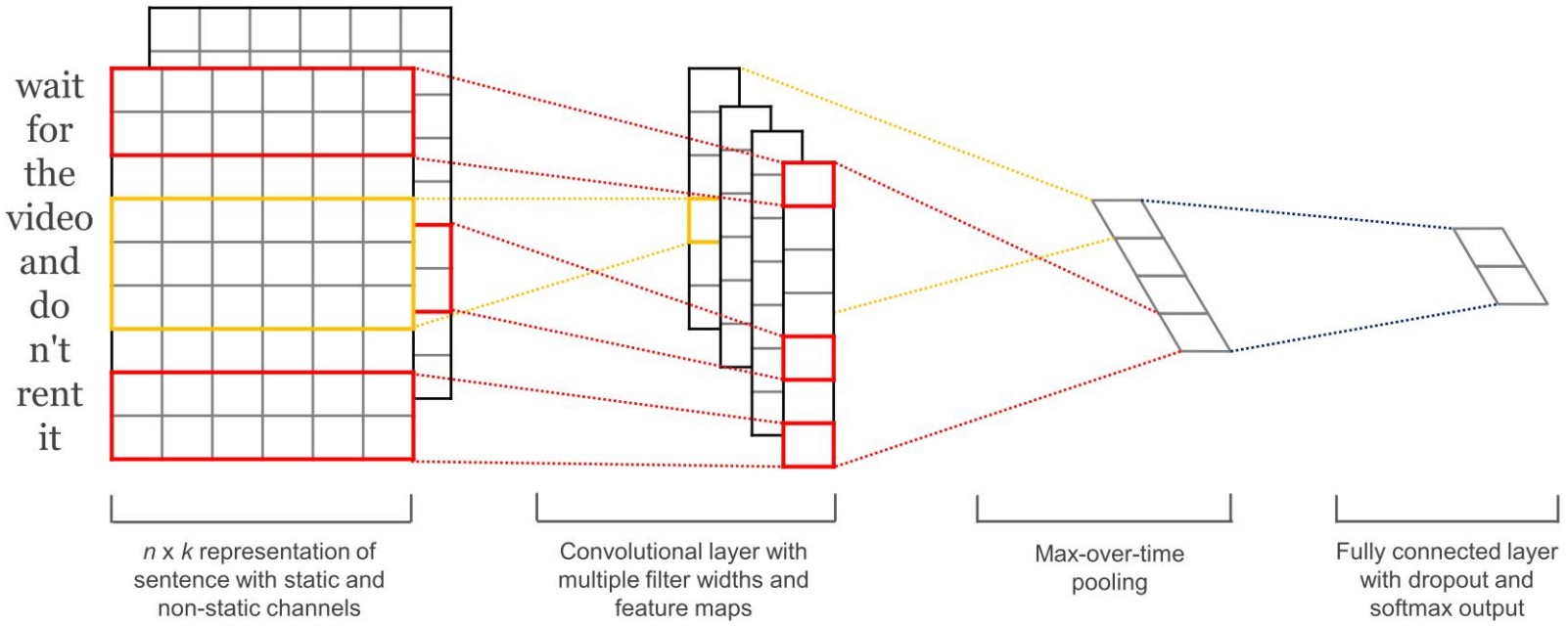
- Search
- Translation
- Question answering
- Document summarization
- Speech recognition
- Entity recognition
- Language modeling
- Sentiment analysis



Tasks overview

https://cloud.google.com/natural-language/

Modern NLP Architectures



http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/

From Frequency to Meaning

tf-idf selects informative terms

DC-9 WITH 55 ABOARD CRASHES; AT LEAST 16 DEAD

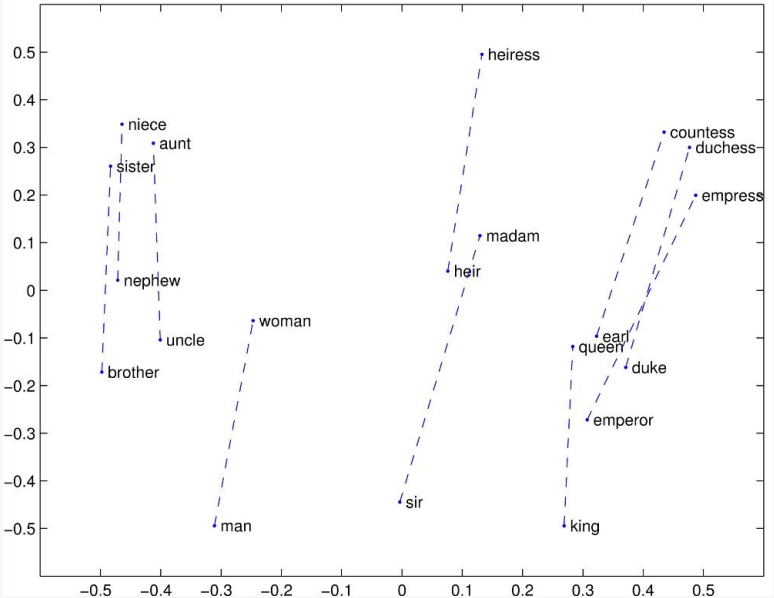
CHARLOTTE, NC, (Reuter)

A USAir DC-9 with 55 people on board crashed and burst into flames during a thunderstorm after missing an approach to Charlotte's international airport Saturday, killing at least 16 people. The flight, which originated in Columbia, South Carolina and was on its final approach, hit a house near the airport runway and caught fire, said Jerry Orr, aviation director at Charlotte-Douglas International Airport. Orr said 16 people were dead, six were missing and presumed dead and 33 were taken to local hospitals. USAir reported 18 dead. Rescue teams fought to save lives inside the wreckage of the plane, which split into three sections on impact at about 6:50 p.m. EDT as the plane was trying to land at Charlotte during heavy storms.

...

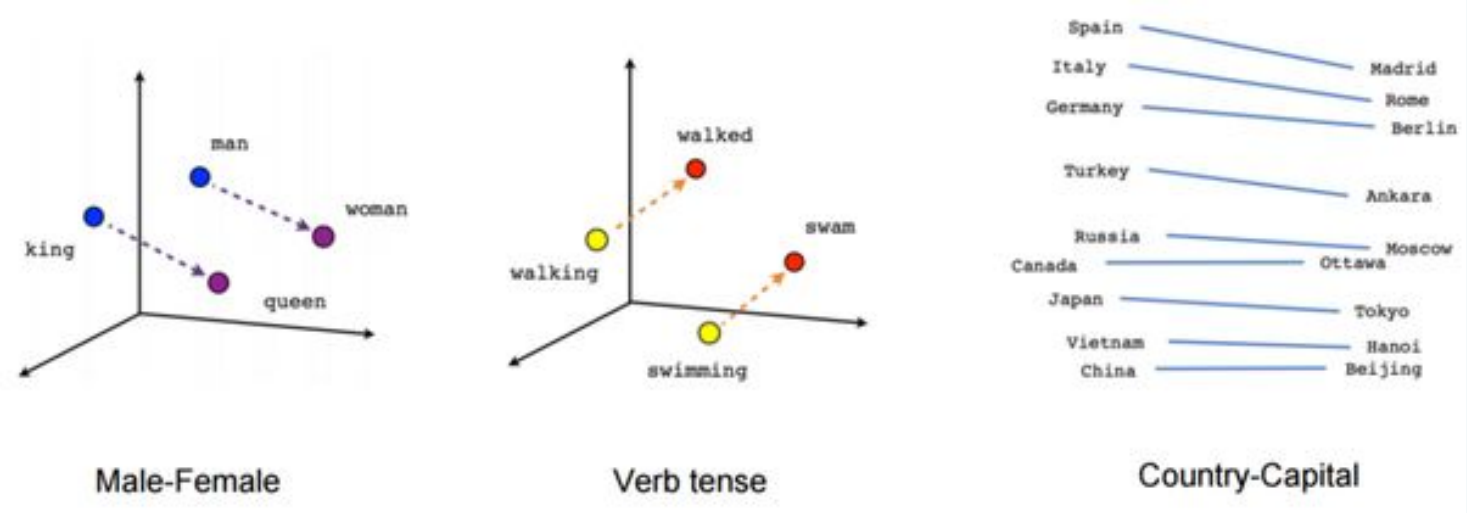
top 15 terms ranked by

frequency	highest idf	tf * idf
32 the	1.00 tdt000077	3.20 orr
16 were	1.00 picknickers	2.81 charlotte
14 said	0.93 sreaming	2.65 payne
12 and	0.93 timmy	2.48 dc
12 to	0.86 6thld	2.24 usair
11 a	0.80 orr	2.00 plane
10 of	0.78 1016	1.93 crash
9 at	0.76 bergen	1.74 bones
9 was	0.75 dripping	1.63 survivors
7 in	0.73 abrams	1.50 dripping
6 on	0.72 0419	1.49 wreckage
6 they	0.69 fuselage	1.35 dead
6 people	0.66 nc	1.29 hospitals
6 had	0.66 thunderstorm	1.27 airport
6 plane	0.66 payne	1.23 55



https://www.youtube.com/watch?v=zvFGNpbAfE, https://nlp.stanford.edu/projects/glove/

Why Word Vectors?



https://www.quora.com/What-is-word-embedding-in-deep-learning

Word Embedding Techniques

Word2vec

- Distributed Representations of Words and Phrases and their Compositionality [T. Mikolov et al.]
- Efficient Estimation of Word Representations in Vector Space [T. Mikolov et al.]

GloVe

- GloVe: Global Vectors for Word Representation [J. Pennington et al.]

FastText

- Enriching Word Vectors with Subword Information [P. Bojanowski et al.]
- Bag of Tricks for Efficient Text Classification [A. Joulin et al.]
- FastText.zip: Compressing Text Classification Models [A. Joulin et al.]

The Distributional Hypothesis

“Words that occur in similar contexts tend to have similar meanings”

The Skip-Gram Model (1)

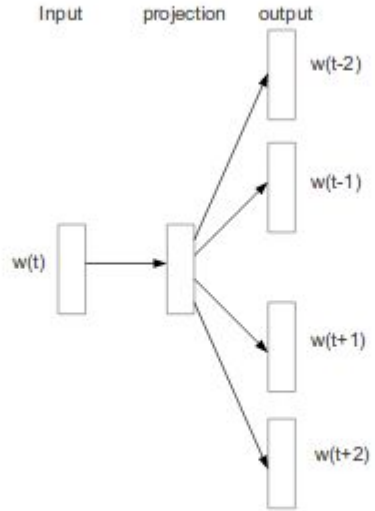


Figure 1: The Skip-gram model architecture. The training objective is to learn word vector representations that are good at predicting the nearby words.

http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

The Skip-Gram Model (2)

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_{t+j}|w_t)$$

$$p(w_O|w_I)=\frac{\exp\left({v'_{w_O}}^{\top}v_{w_I}\right)}{\sum_{w=1}^{W}\exp\left({v'_w}^{\top}v_{w_I}\right)}$$

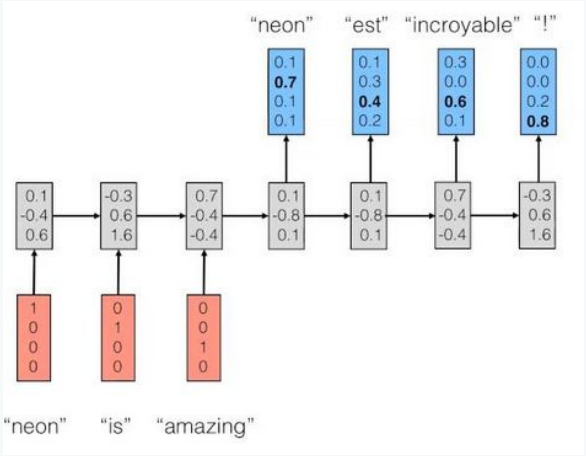
http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

The Skip-Gram Model (3)

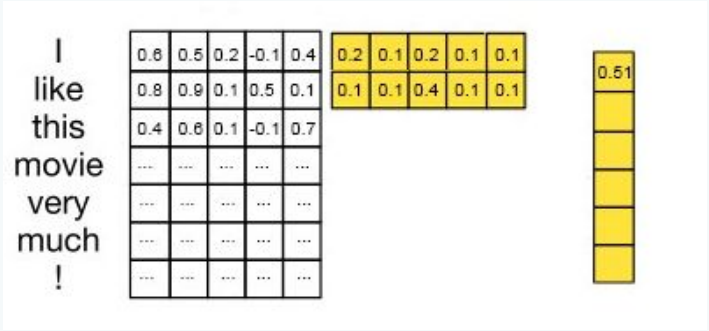
https://projector.tensorflow.org/

Deep Learning for Text

RNNs



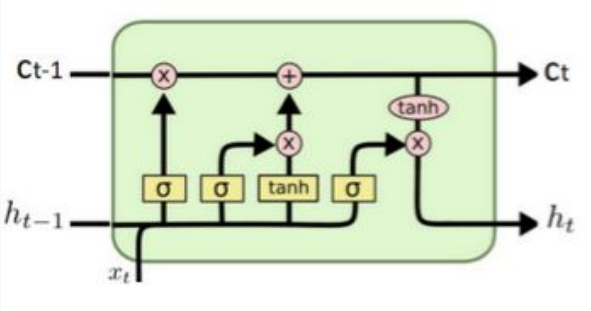
CNNs



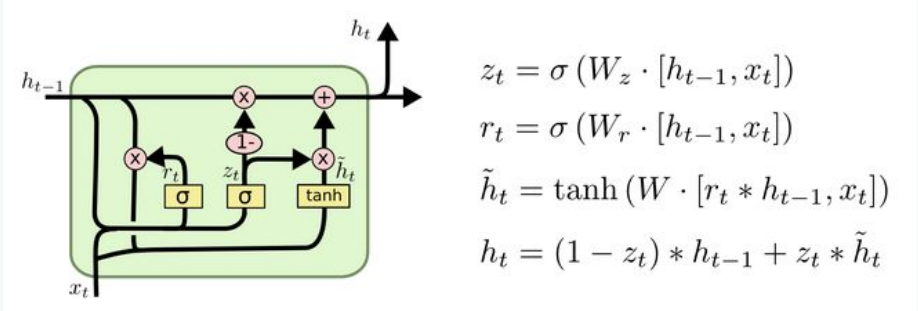
https://towardsdatascience.com/understanding-how-convolutional-neural-network-cnn-perform-text-classification-with-word-d2ee64b9dd0b,
https://www.youtube.com/watch?v=Ukgii7Yd_cU

RNNs for Text

LSTMs

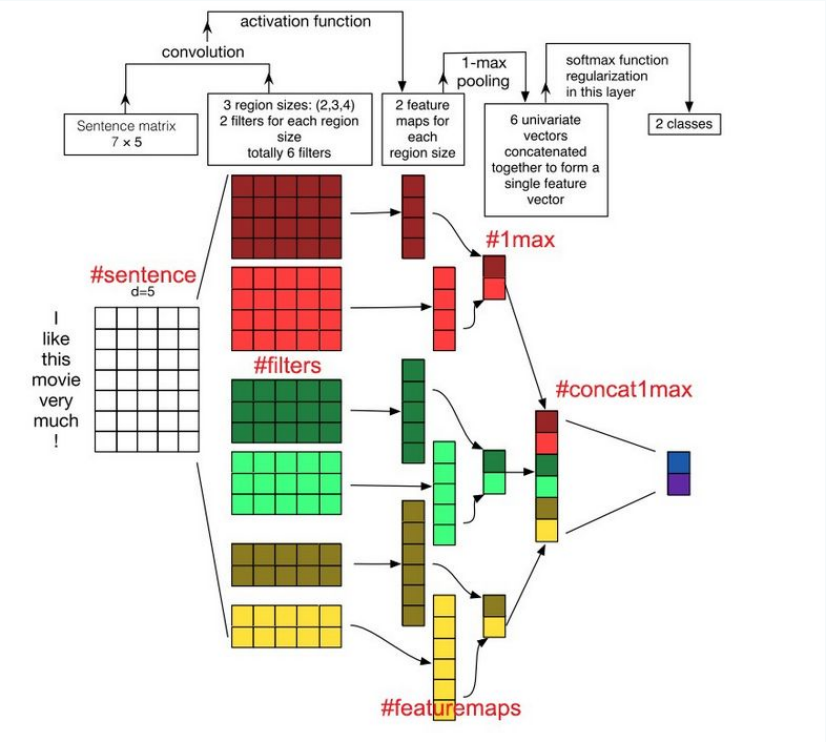


GRUs



http://colah.github.io/posts/2015-08-Understanding-LSTMs

CNNs for Text



https://towardsdatascience.com/understanding-how-convolutional-neural-network-cnn-perform-text-classification-with-word-d2ee64b9dd0b

RNN or CNN ?

RNN

- Sequential architecture
- Efficient modelling of context dependencies
- Translation, language modelling

CNN

- Hierarchical architecture (faster forward pass and backward pass)
- Efficient key phrase recognition
- Sentiment detection, named entity recognition

-> How important is it to semantically understand the whole sequence ?

https://arxiv.org/pdf/1702.01923.pdf

Sample Code in Keras (1)

```
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence

max_text_length = 100
vocab_size = 10000
X_train = sequence.pad_sequences(X_train, maxlen=max_length)
X_test = sequence.pad_sequences(X_test, maxlen=max_length)

word_vector_dim = 100
model = Sequential()
model.add(Embedding(vocab_size, word_vector_dim, input_length=max_text_length))
model.add(LSTM(64))
model.add(Dense(32, activation='relu'))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
model.fit(X_train, y_train, validation_data=(X_test, y_test), epochs=20, batch_size=64)
```

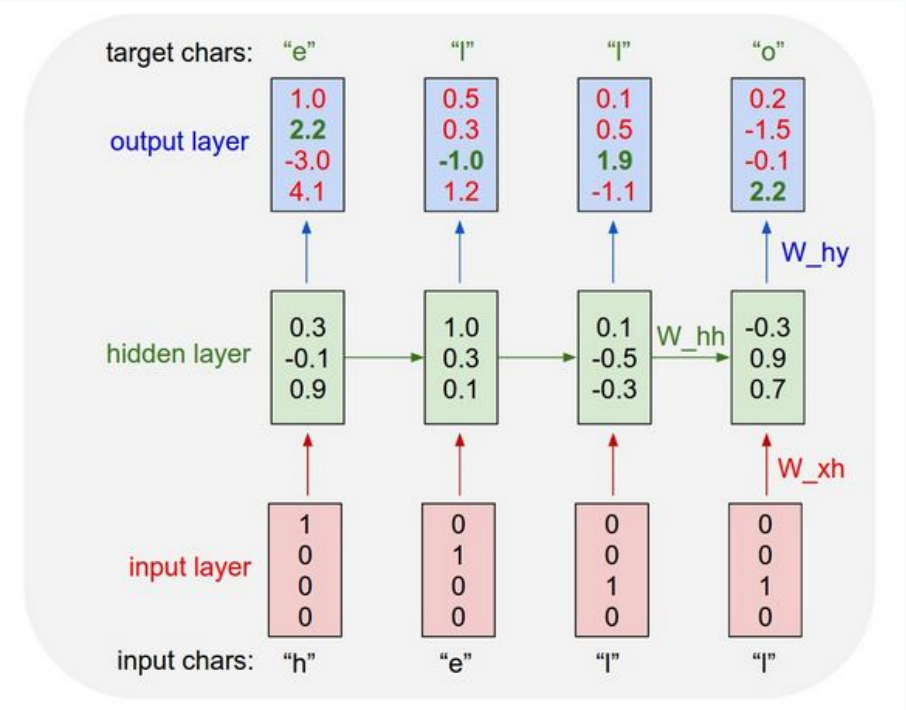
Sample Code in Keras (2)

```
from keras.models import Sequential
from keras.layers import Conv1D, MaxPooling1D, Dense
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence

max_text_length = 100
vocab_size = 10000
X_train = sequence.pad_sequences(X_train, maxlen=max_length)
X_test = sequence.pad_sequences(X_test, maxlen=max_length)

word_vector_dim = 100
model = Sequential()
model.add(Embedding(vocab_size, word_vector_dim, input_length=max_text_length))
model.add(Conv1D(128, 5, activation='relu'))
model.add(MaxPooling1D(5))
model.add(Conv1D(128, 5, activation='relu'))
model.add(MaxPooling1D(8))
model.add(Flatten())
model.add(Dense(32, activation='relu'))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
model.fit(X_train, y_train, validation_data=(X_test, y_test), epochs=20, batch_size=64)
```

A Practical Example

http://reveal.riminder.net/

How to Generate Text ?



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Shakespeare

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Simpsons

```
moe_szyslak: uh, hey, how ya doin'?
homer_simpson: i was just tellin' all the bad news to not fail.
homer_simpson: yeah, me, you'd treat her right.(regretful) as a
little one, i think of my treasure.
moe_szyslak: oh guys, it was horrible.
moe_szyslak: ya bunch of ungrateful ingrates! ya--
carl_carlson: you got this?
moe_szyslak: no, no, no. not a little girl will be in the air.
moe_szyslak: yeah, you don't even have a beer?
homer_simpson:(sunk) i dunno.


moe_szyslak: sure.
homer_simpson: number a mean, or a deal. i didn't mean that.(to
home) there's a thing i call my man. i am not an angel!
moe_szyslak: well, i guess the world's smallest violin. and you
can just waltz off the
homer_simpson: to a...


moe_szyslak:(cutting him off) too late, or is it.
homer_simpson: moe, i could forget ya.
```

https://medium.com/@harshitwas/making-your-own-simpsons-tv-script-7f791581bf03

Sunspring: Sci-Fi Movie written by AI

http://cutprintfilm.com/features/vimeo-short-film-week-past-inside-present/

Conclusion

One cool thing

https://www.youtube.com/watch?v=vZFNzwv61Fk
https://www.youtube.com/watch?v=Eo07BAsyQ24

Point of Contact

Paul Coursaux
Lead AI Researcher

paul.coursaux@riminder.net