



达伦
دارن
ダレン

库克
كوك
クック

DARREN COOK

DIRECTOR

dc@QQTREND.com

O'REILLY®



Data Push Apps with HTML5 SSE

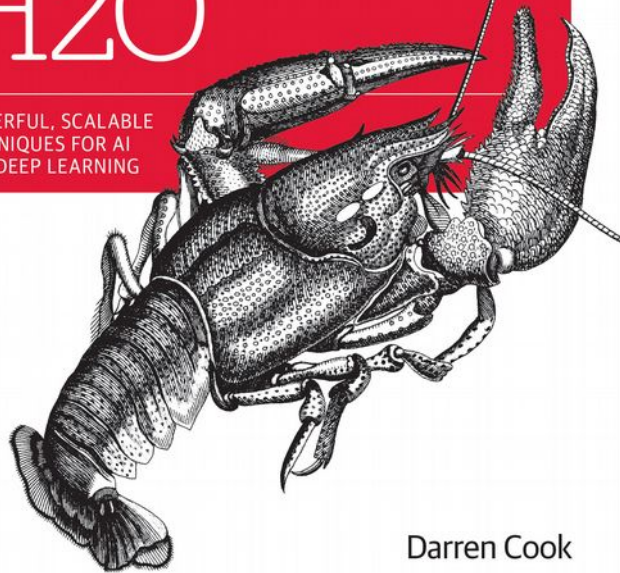
PRAGMATIC SOLUTIONS FOR REAL-WORLD CLIENTS

Darren Cook

O'REILLY®

Practical Machine Learning with H2O

POWERFUL, SCALABLE
TECHNIQUES FOR AI
AND DEEP LEARNING



Darren Cook



Company Values & Vision

Contact

Company Values & Vision

Here at QQ Trend, the following five principles have guided us:

- Freedom
- Global Mindset
- Data is our horizontal
- Leave It Better Than We Found It
- The Q is for Quality

<https://blog.qqtrend.com/values>

At the very heart of our values is the belief that open source and open data are Good Things

Search depth 14
 Highest Value 37
 Lowest Value 7
 Total Nodes
 Time Spent (s)
 Mean 7.86
 S.D. 3.14

Data Set 0124
 Release 03 Mode
 Version 1.0.17

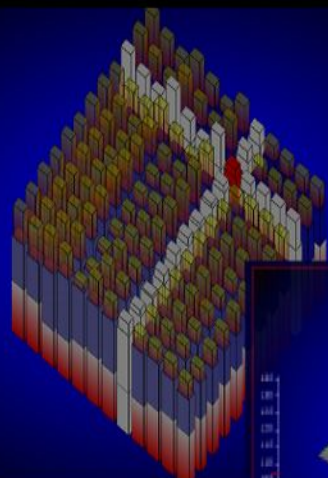


チャート デモ

Version 2014-09-10
 Copyright © 2014, 2015
 All Rights Reserved

Year	Q1	Q2	Q3	Q4
2004	100.0	100.0	100.0	100.0
2005	100.0	100.0	100.0	100.0
2006	100.0	100.0	100.0	100.0
2007	100.0	100.0	100.0	100.0
2008	100.0	100.0	100.0	100.0
2009	100.0	100.0	100.0	100.0
2010	100.0	100.0	100.0	100.0
2011	100.0	100.0	100.0	100.0
2012	100.0	100.0	100.0	100.0
2013	100.0	100.0	100.0	100.0
2014	100.0	100.0	100.0	100.0
2015	100.0	100.0	100.0	100.0

3Dチャート
 2Dチャート
 4半期グラフ

English Auto Backtest
 日本語
 日本語
 日本語

100% 100% 100% 100%

QQ Trend Ltd.

- Investors
- Clients
- Working with us

(<https://blog.qqtrend.com/values/>)



- What is hard about MT
- Why multiple parts/data
- Looking at the parts
- Open-source/data projects



- For professional translators
- NMT, word embeddings, latest tokenizers, latest dictionaries, ...
- Enforce style, terminology, ...



SENTIENCE

HUMAN INSPIRATION. MACHINE INTELLIGENCE.

050-5532-6522

INBOX@SENTIENCE.JP

- Ambitious...
- Trados at the high-end
- Google Translate et al at the low-end



SENTIENCE

HUMAN INSPIRATION. MACHINE INTELLIGENCE.

050-5532-6522

INBOX@SENTIENCE.JP

Machine Translation Is Hard



coup de foudre

coup ⚡ → blow, hit, bang

de ⚡ → of, to, from

foudre ⚡ → lightning



1. Read/hear
2. Internalize, using context
3. Write/speak



Mrs.Suzuki is a great teacher!

鈴木さんはいい先生だ！

Mr.Suzuki is a great teacher!



- Gisting
- Perfectly accurate translation
- Professional translation



- Gisting
- Perfectly accurate translation
- Professional translation

“How accurately does the candidate text convey the semantics of the source text?”



自己免疫疾患では、注射剤ではあるがエンティビオ（武田薬品、上市）やエトロリズマブ（ロシュ、フェーズIII）といった抗インテグリン抗体も競合になりうる。



For autoimmune diseases, although it is an injection, anti-integrin antibodies such as Entivio (Takeda Pharmaceutical Co., Ltd.) and Ethenrolizumab (Roche, Phase III) can also compete.



In the field of autoimmune diseases, injectable anti-integrin antibodies such as Takeda's Entyvio (currently marketed) and Roche's etrolizumab (Phase 3) are potential competing drugs.



自己免疫疾患では、注射剤ではあるがエンティビオ（武田薬品、上市）やエトロリズマブ（ロシュ、フェーズⅢ）といった抗インテグリン抗体も競合になりうる。

For autoimmune diseases, although it is an injection, anti-integrin antibodies such as Entivio (Takeda Pharmaceutical Co., Ltd.) and Etorolizumab (Roche, Phase III) can also compete.

In the field of autoimmune diseases, injectable anti-integrin antibodies such as Takeda's Entivio (currently marketed) and Roche's etrolizumab (Phase 3) are potential competing drugs.

- Readability
- Accuracy
- Usefulness



<https://www.theatlantic.com/technology/archive/2018/01/the-shalowness-of-google-translate/551570/>

(en-fr, de-en, zh-en)

Douglas Hofstadter, a professor of cognitive science and comparative literature at Indiana University at Bloomington

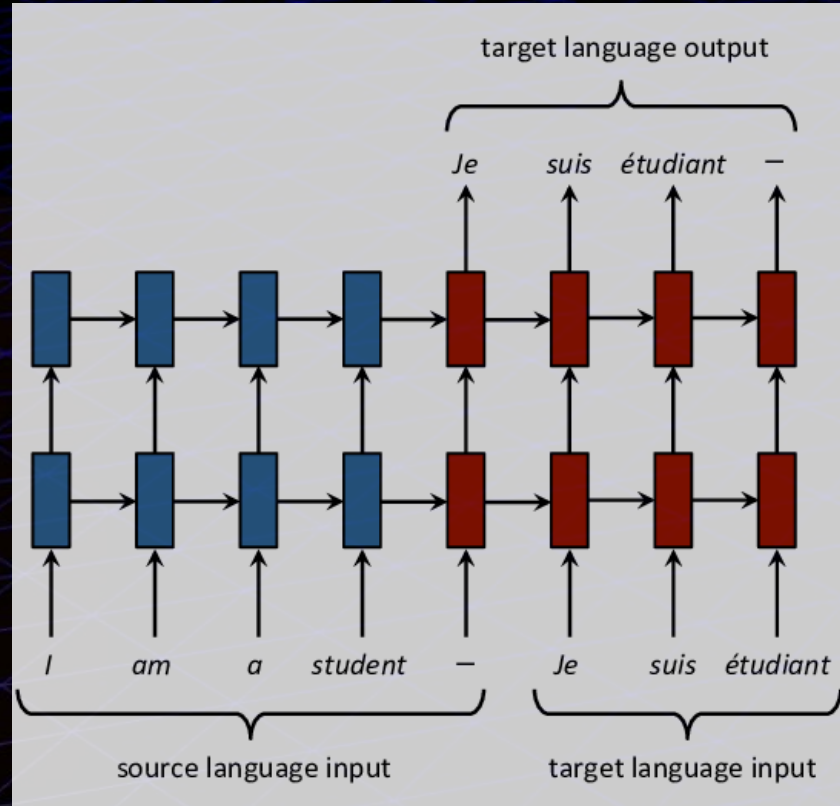


It's almost irresistible for people to presume that a piece of software that deals so fluently with words must surely know what they mean.

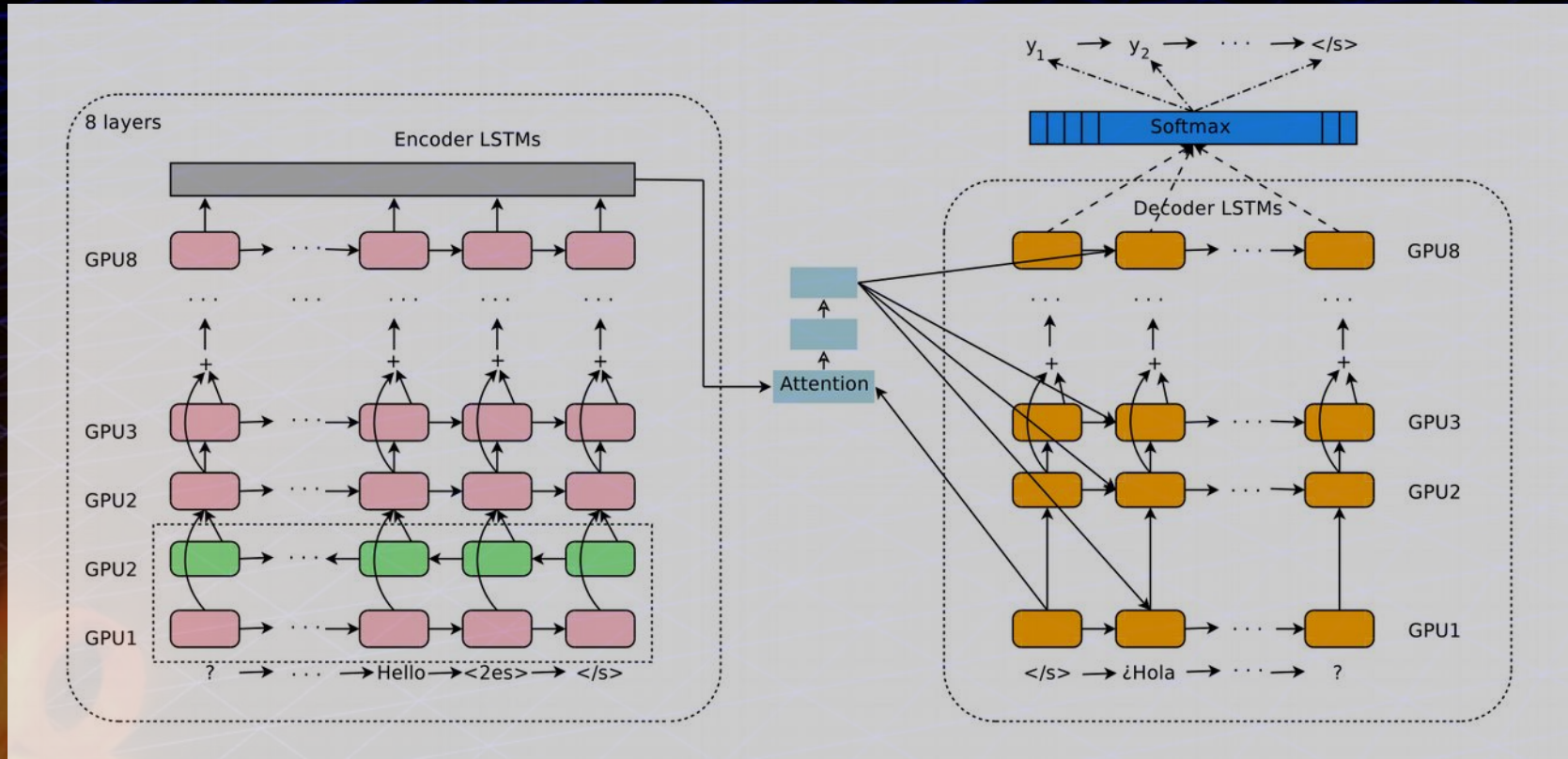
(Douglas Hofstadter)



NMT: one sentence at a time



NMT: slightly less simplified



Source: google paper, arxiv.org/pdf/1609.08144

One more motivating example...

鈴木社長によると ...

Suzuki CEO according to...



One more motivating example...

鈴木社長によると ...

According to Mr. Hiromi Suzuki, CEO, ...



In other words...

Context



Context

Context



- In the cloud, or offline?
- Mobile or desktop or web app?
- Cross-platform?
- Computer language?
- Essential tools/libraries?
- Licenses?



- In the cloud, or offline?
- Mobile or desktop or web app?
- Cross-platform? **Electron (Win/Mac/Linux)**
- Computer language? **JavaScript**
- Essential tools/libraries?  
- Licenses? **MIT/Apache, CC-BY**



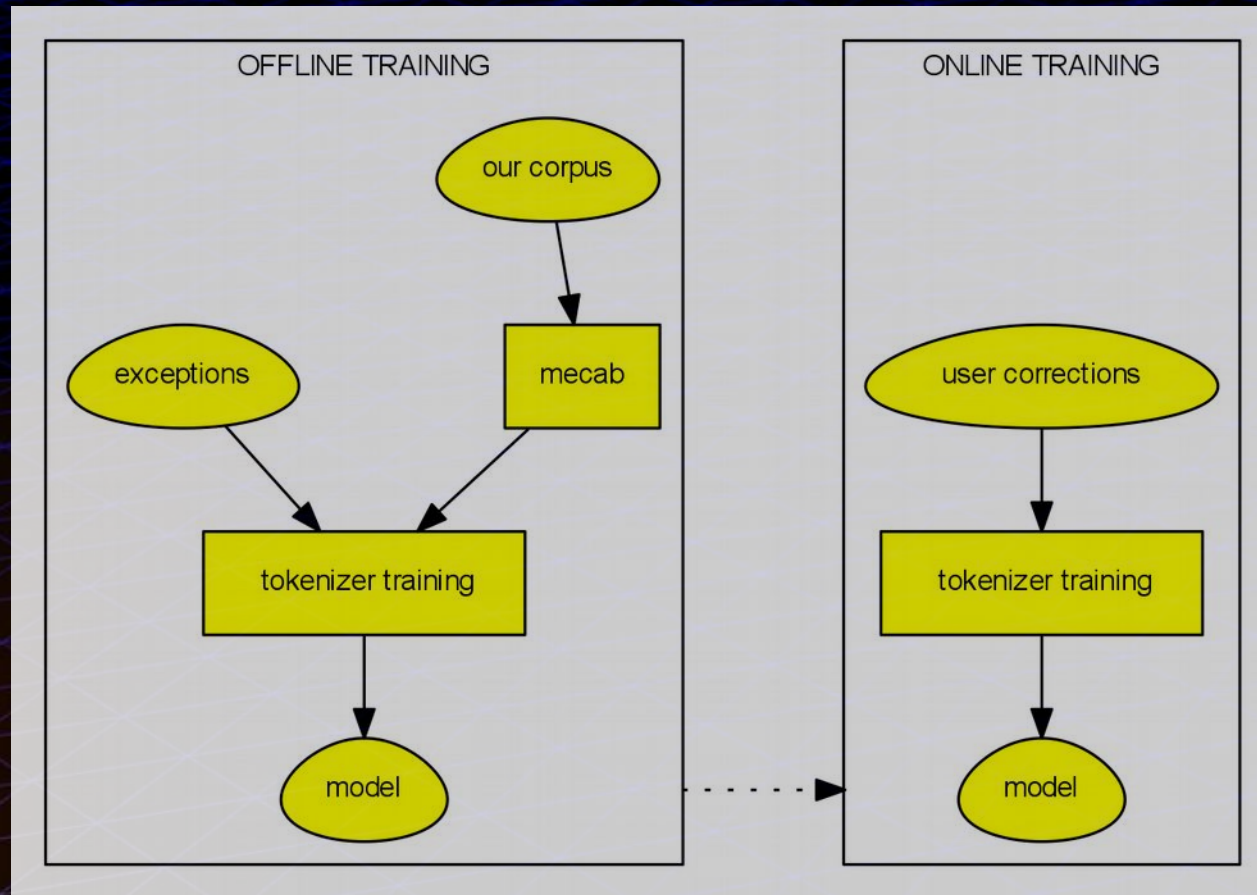
自己免疫疾患では、注射剤ではあるがエンティビオ（武田薬品、上市）やエトロリズマブ（ロシュ、フェーズIII）といった抗インテグリン抗体も競合になりうる。



自己 [N-nc] | 免疫 [N-nc] | 疾患 [N-nc] | で [P-k] |
は [P-rj] | 、 [M-c] | 注射 [N-nc] | 剤 [Q-n] |
で [P-k] | は [P-rj] | ある [V-dp] | が [P-sj] |
エンティビオ [N-nc] | ([M-op] | 武田 [N-pn] | 薬 [Q-n] |
品 [Q-n] | 、 [M-c] | 上市 [N-nc] |) [M-cp] | や [P-fj] |
エトロリズマブ [N-nc] | ([M-op] | ロシュ [N-pn] |
、 [M-c] | フェーズ [N-nc] | III [N-nc] |) [M-cp] |
と [P-k] | いっ [V-c] | た [X] | 抗 [P] |
インテグリン [N-nc] | 抗体 [N-nc] | も [P-rj] |
競合 [N-nc] | に [P-k] | なり [V-dp] | うる [V-dp] | 。 [M-p]

Using: <https://github.com/rakuten-nlp/rakutenma>





Also Q7/Q7K, other normalization



Stop words, word embeddings



























Stop words

- This is a problem with English.
- This is a problem with the English



This is a problem with English.



Meet Met Met		ride rode ridden		Go Went gone	
Hide Hid Hidden		Spend Spent spent		Take Took Taken	
Get Got Got		Show Showed shown		Run Ran Run	
Hit Hit Hit		Fight Fought Fought		Swim Swam swum	
Say Said Said		Sit Sat Sat		Build Built Built	
Lose Lost Lost		Draw Drew Drawn		Sing Sang sung	
Grow Grew Grown		Give Gave Given		Lend Lent Lent	
Feel Felt Felt		Feed Fed Fed		Pay Paid Paid	



This is a problem with
the English.

Text Generation

Cat sat mat.



Word Embeddings

Consistency matters

So, we generate our own



LSTMs



LSTMs

- 27 characters?
- 50K words? More? Less?
- 8000 chars for Chinese?
- 2100 chars for Japanese?



Unknown words

- UNK
- UNK-A, UNK-B, ...
- Q-UNK-h, ...
- Abbreviations
- Proper Nouns



LSTMs in JavaScript

TensorFire

<https://tenso.rs/>



Using TensorFire

1. Prepare our data
2. Train an LSTM model in Keras(tensorflow)
3. Conversion script (TensorFlow → TensorFire)
4. Load that binary model in our application
5. Make predictions from JS



LSTMs in JavaScript

TensorFire

<https://tenso.rs/>

TensorFlow JS

<https://js.tensorflow.org/>



Data Sources

Dictionaries



Semantic Networks



Disambiguation

吉田社長

If Sony: Kenichiro Yoshida

If TV Asahi: Shinichi Yoshida



Open Source/Data Projects

(Always liberal licenses: MIT or CC-BY)

- qqNLP
- Multilingual semantic network
- Company database
- Bad examples





达伦
دارن
ダレン

库克
كوك
クック

DARREN COOK

DIRECTOR

dc@QQTREND.com