Practical NLP

Tal Perry, CEO LightTag

Light TAG

Goals

- To become familiar with basic and advanced tools for NLP
- To know when and how to use which
- To gain practical experience with those tools

This is a hands on course

You will read documentation and write code.
I won't talk much

"People have been studying calculus for 300 years. Coming to class is stupid. Go read a book and ask me questions in my office"

- *My Calculus professor, Shiri Arstein*

Light TAG

This is a hands on course

You will read documentation and write code.
I won't talk much

There's abundant content on the internet for NLP.

This course is an opportunity for you to work through the hard technical details.

By the end, you should be able to lookup the cutting edge techniques and

1. Decide if you should use them
2. Implement and adjust them to your needs

One Problem

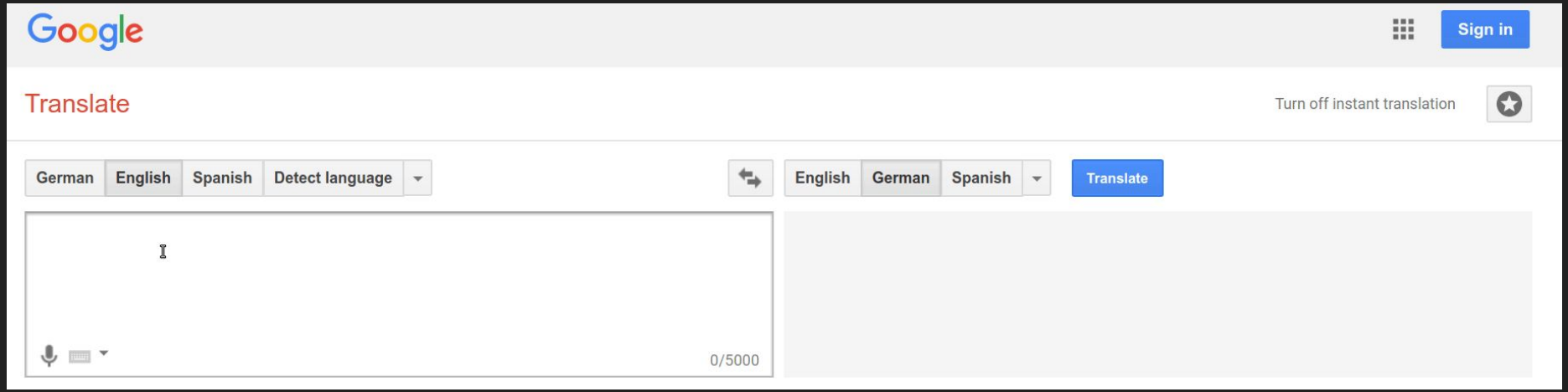
vierzehnmillionfünfhundreteinundzwanzigtausendvierhundretfünfzehn

We'll spend the course looking at a single problem:

Given a number as a word in German, convert it to it's numerical form

This focus allows us to isolate the phenomena that's important from the boilerplate around it.

One Problem



LightTAG

The Plan

Day 1

“Classical Methods”

- **Regular expressions**
- **Tokenization**
- **The bag of words model**
- **Linear Models for text classification**

The Plan

Day 2

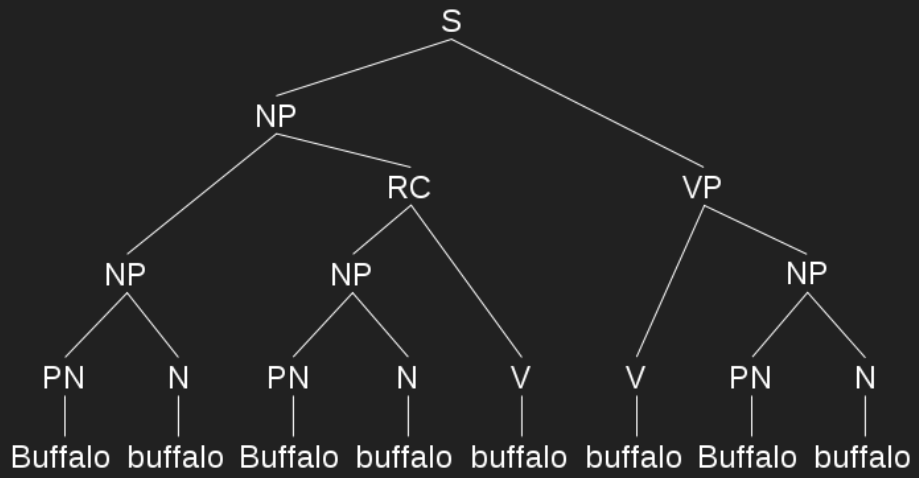
“Deep Learning for NLP”

- **Distributed Representations**
- **Handling Sequences with RNNs**
- **Getting Text into Tensorflow**

A fun example on why language is hard

What has 4 letters, sometimes 9 letters, but never has 5 letters.

A fun example on why language is hard



Light TAG

A Unified Theory of Inference for Text Understanding

By

Peter Norvig

B.S. (Brown University) 1976

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

in the

GRADUATE DIVISION

OF THE

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved: 11/24/86
Chairman Date

..... 11/25/86

..... 11/25/86

What we can't do

People are very good at interpreting texts and making inferences. They generally do not notice when the text is under-specified and they have to make inferences to resolve ambiguities, or to gain a fuller understanding of the text. As an example, consider the following text, excerpted from a book of fairy tales [9]. It will be referred to as text (1).

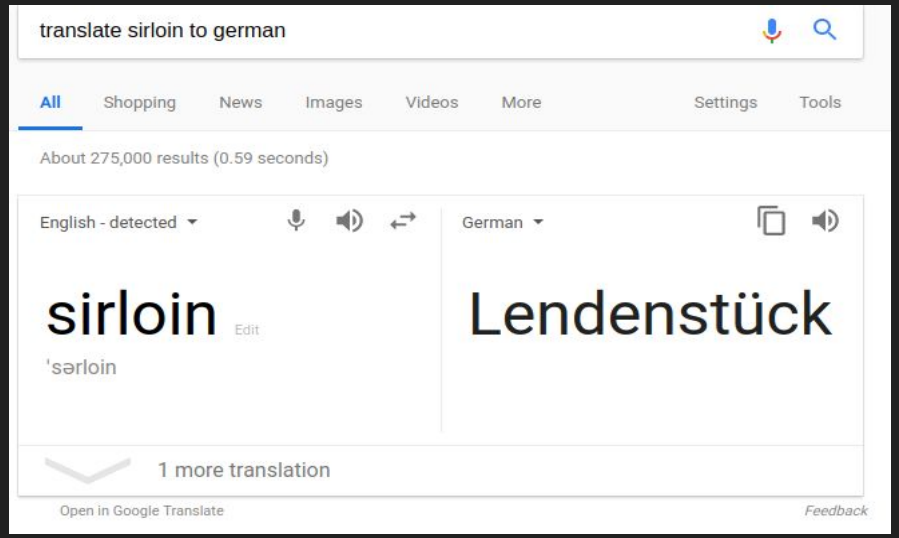
> In a poor fishing village built on an island not far from the coast of China, a young boy named Chang Lee lived with his widowed mother. Every day, little Chang bravely set off with his net, hoping to catch a few fish from the sea, which they could sell and have a little money to buy bread.
> . . .

A reader of text (1) should be able to make inferences like these:

(2a) There is a sea which is used by the villagers for fishing, surrounds the island, and forms the coast of China.
(2b) Chang intends to trap fish in his net, which is a fishing net.
(2c) The word *which* in *which they could sell* refers to the fish.
(2d) The word *they* in *they could sell* refers to Chang and his mother.

What we can't do

We don't know how to "infer" *obvious* information from text



What we can do

- Translation
- Entity Recognition
- Sentiment Analysis
- Imputation
- Dependency Parsing
- Why is this hard ?
  - “The rules” are ambiguous
  - People don’t follow them
  - All data is dirty

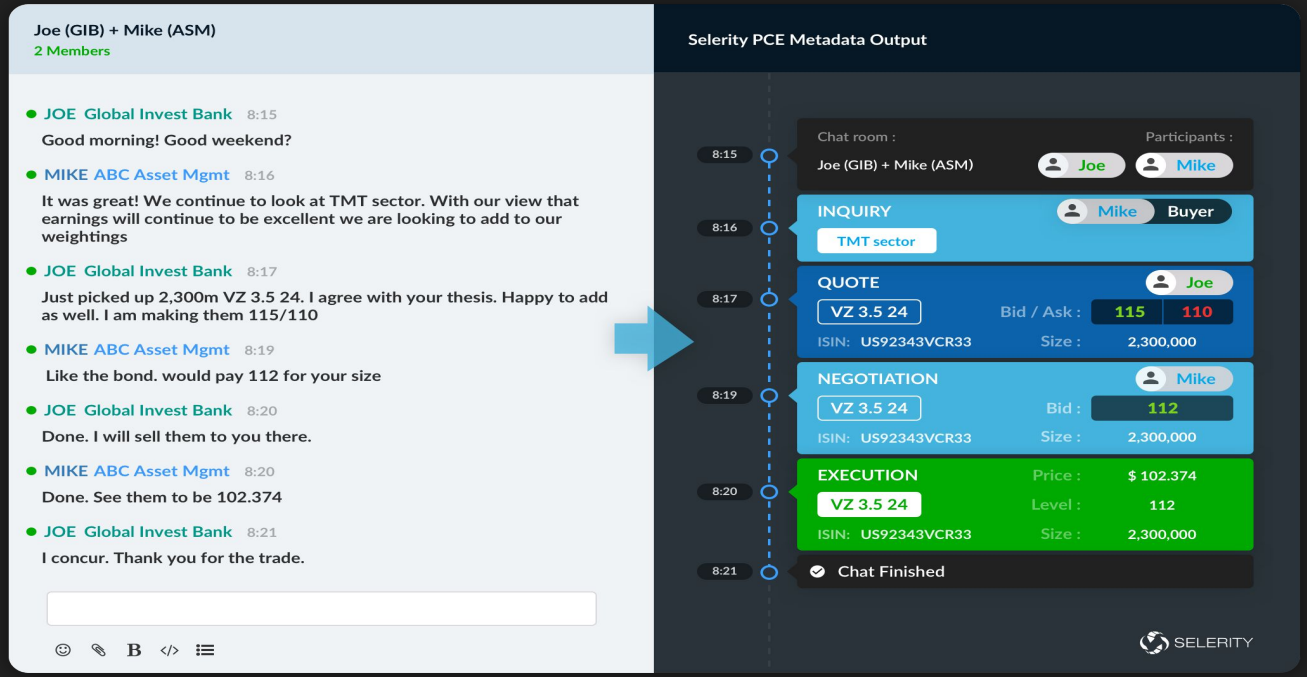
Light TAG

What we want to do in industry

- Find information and organize it in useful ways

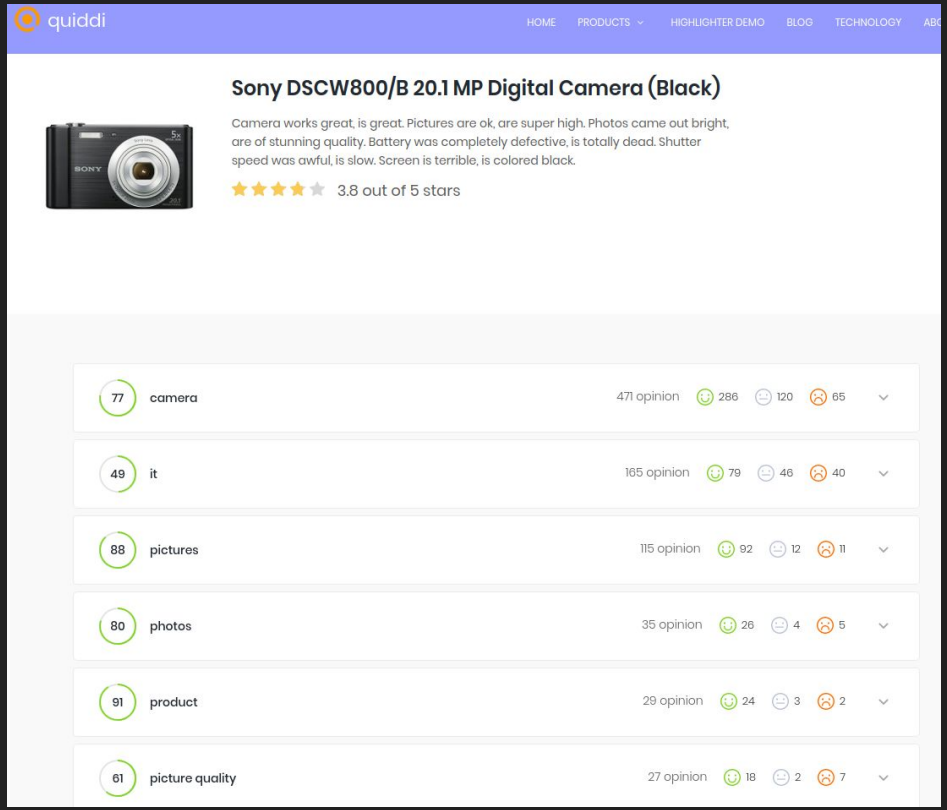
LightTAG

NLP In Industry



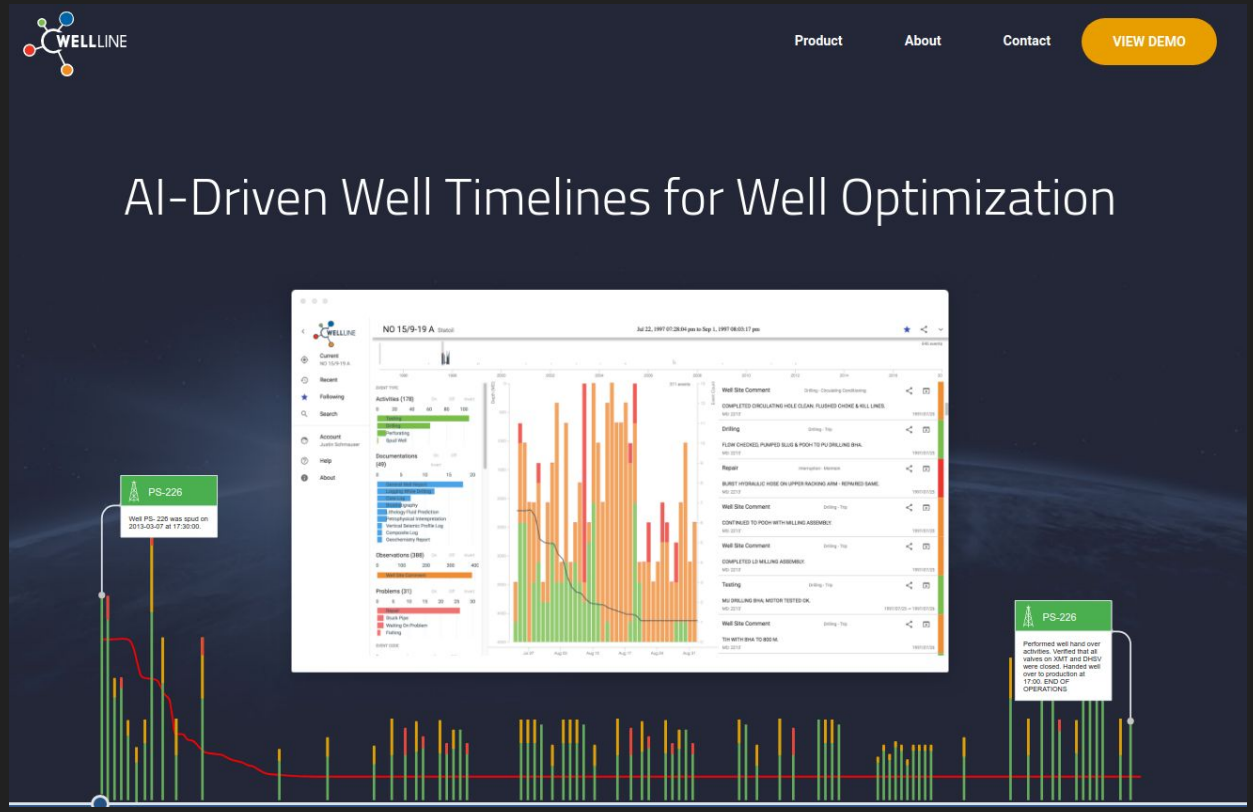
Light TAG

NLP In Industry



LightTAG

NLP In Industry



LightTAG

Self Study Resources

- CS244D/N Stanfords Deep NLP course
- Speech and Language Processing - Dan Jurafsky and James Martin
- A Primer on Neural Network Models for Natural Language Processing -Yoav Goldberg