

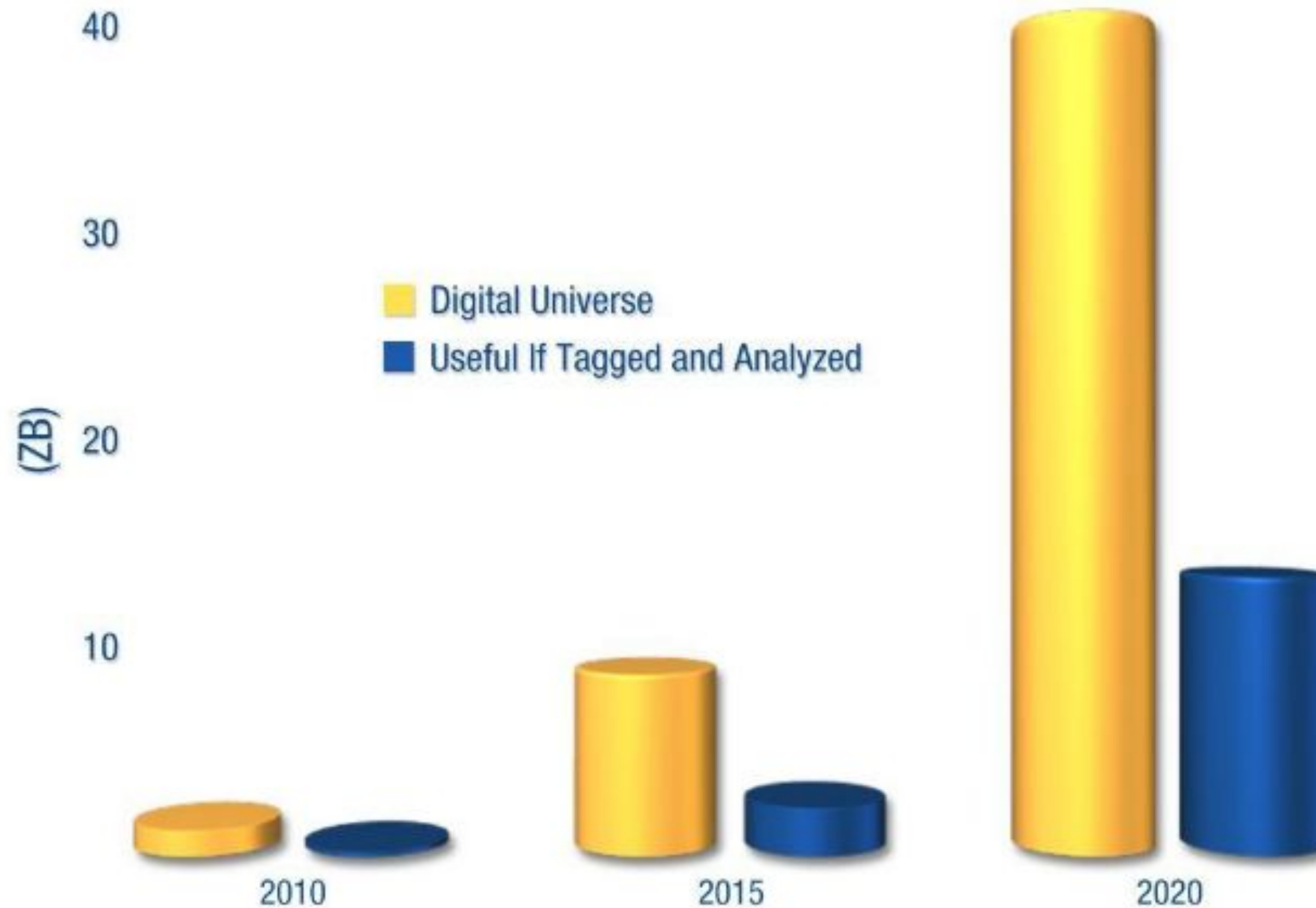
Análise de Mídias Sociais e Mineração de Texto

Mineração de Texto

Laura de Oliveira F. Moraes

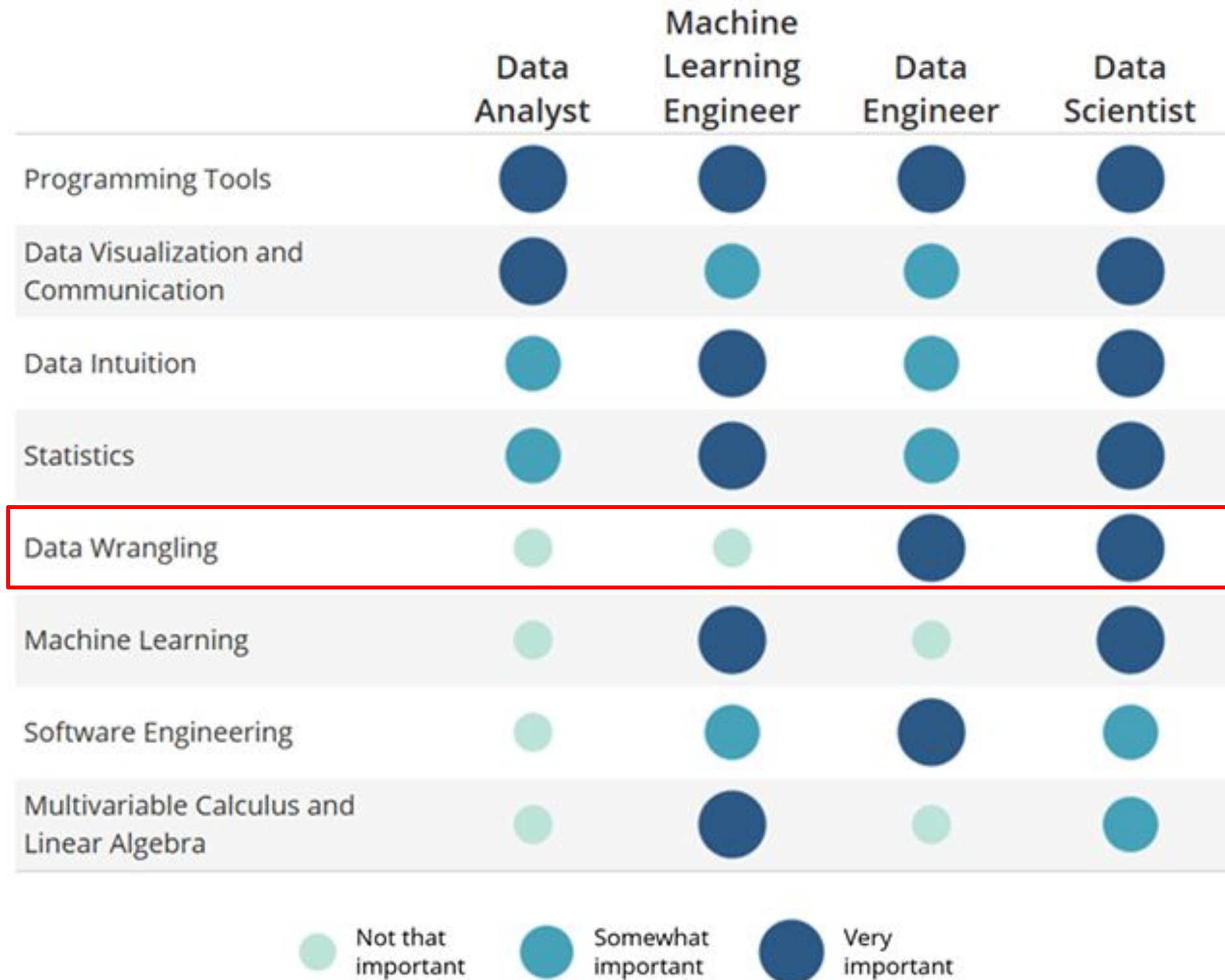
Por que?

Opportunity for Big Data



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

Por que?



NLP são métodos para computadores analisarem, entenderem e obterem significados para a linguagem humana de uma maneira inteligente e útil

NLP são **métodos** para computadores **analisarem, entenderem e obterem significados** para a linguagem humana de uma **maneira inteligente e útil**

Compreensão
da Linguagem

Representar a linguagem

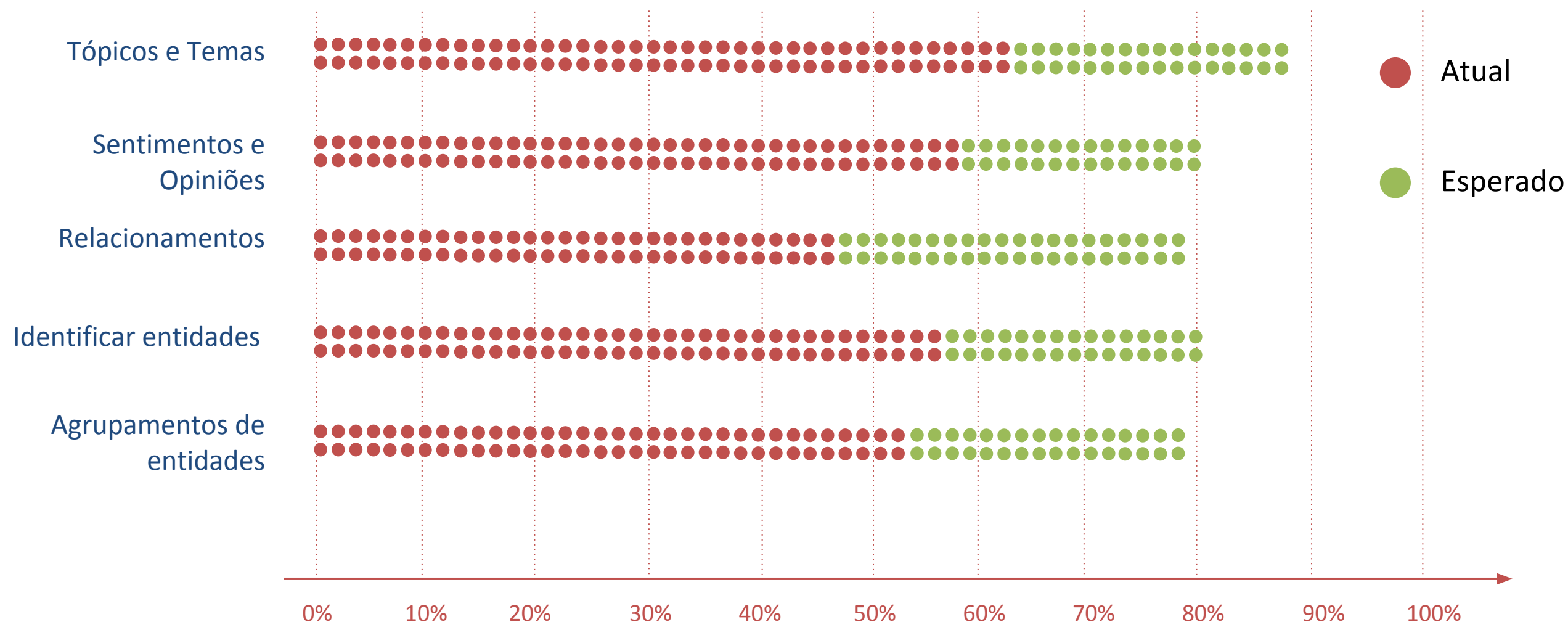
Analisar diferentes **aspectos**
da linguagem

Geração da
Linguagem

Produzir conteúdo
significativo

Entre outras aplicações: **Bots**

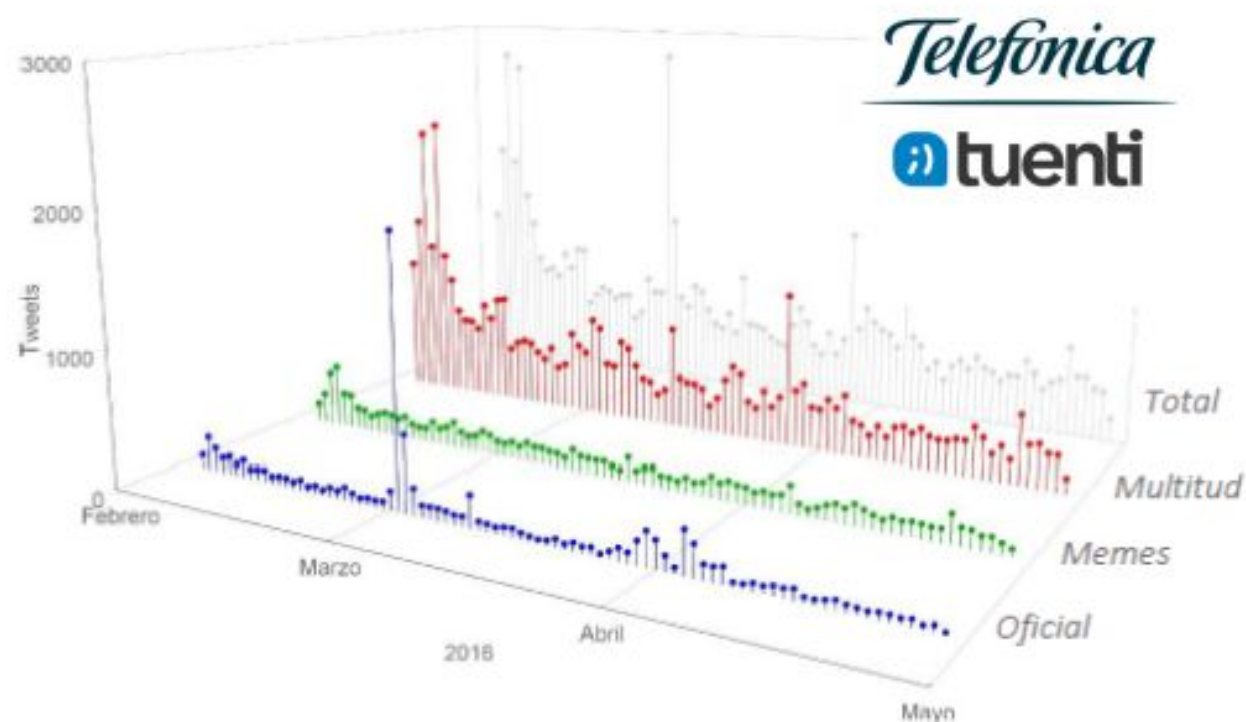
Sobre análises de conteúdo textual, o que você necessita, ou espera ter como necessidade?



Painel Vila Olímpica



Prefeitura Vila Olímpica



Common NLP Tasks



Easy

- Chunking
- Part-of-Speech Tagging
- Named Entity Recognition
- Spam Detection
- Thesaurus



Medium


- Syntactic Parsing
- Word Sense Disambiguation
- Sentiment Analysis
- Topic Modeling
- Information Retrieval



Hard

- Machine Translation
- Text Generation
- Automatic Summarization
- Question Answering
- Conversational Interfaces

Dados não-estruturados, enriquecidos



TOP STORIES MEDIA CENTER PROGRAM LE

GERMANY WORLD BUSINESS SCI-TECH ENVIRONMENT CUL

TOP STORIES / BUSINESS

TECHNOLOGY


Google buys stake in SpaceX to be Internet from outer space

Google has secured a 10 percent stake in the private space exploration company SpaceX along with another investor. Reports suggest the money is aimed at hooking up far-off areas with Internet access from outer space.

The most remote, barren corners of the globe may soon have Internet access.

SpaceX the private space exploration company backed Tesla Motors and Elon Musk did Tuesday it had raised \$1 billion (860 million euros) from two new investors Google and Fidelity Investments.

SpaceX




Space Exploration Technologies Corp logo

Type	Private
Industry	Aerospace
Founded	2002
Founder	Elon Musk
Headquarters	Hawthorne, California, USA
Key people	Elon Musk (CEO and Chief Designer) Gwynne Shotwell (President and COO) Tom Mueller (VP of Propulsion)
Services	Orbital rocket launch
Number of employees	3,000+ (Jan 2015)[citation needed]
Website	SpaceX.com

Footnotes / references [1][2][3]

Google Inc.



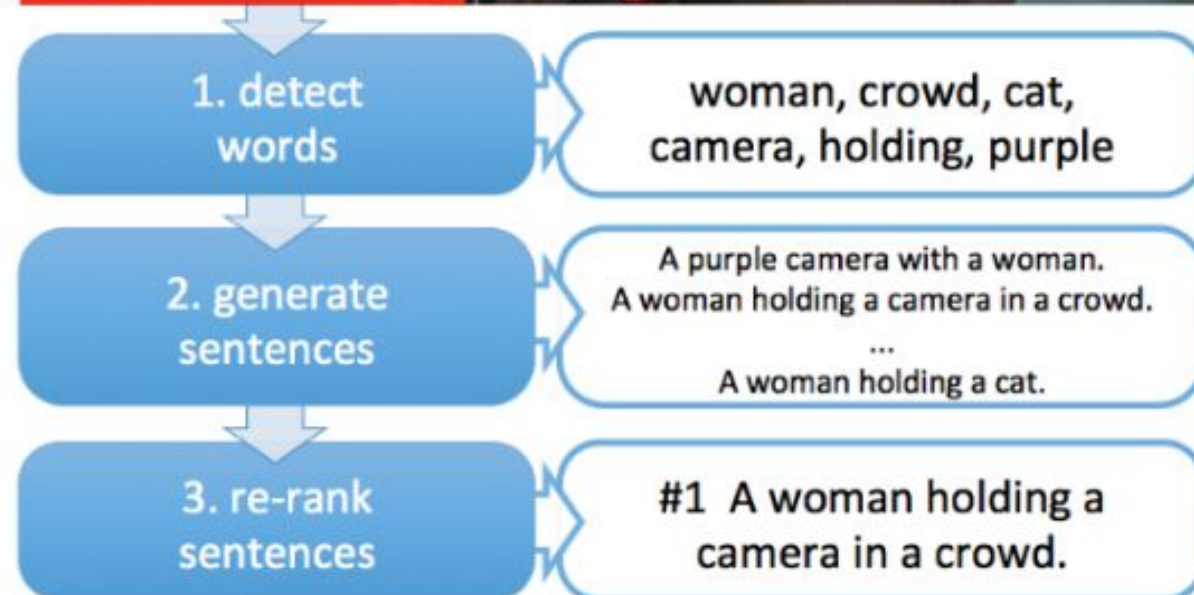
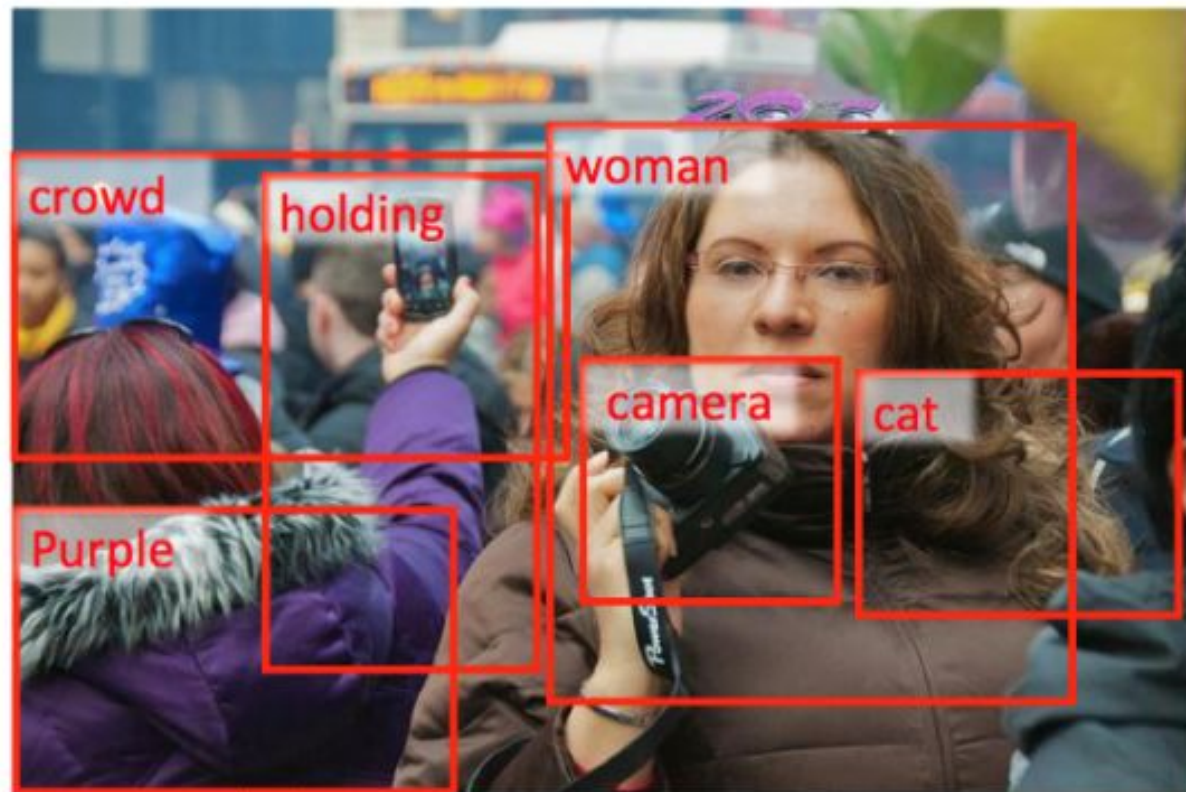
Type	Public
Traded as	Class A: NASDAQ: GOOGL Class B supervoting: unlisted Class C nonvoting: NASDAQ: GOOG NASDAQ-100 Components (GOOGL and GOOG) S&P 500 Components (GOOGL and GOOG)
Industry	Internet Computer software Telecoms equipment
Founded	September 4, 1998; 16 years ago Menlo Park, California[1][2]
Founder	Larry Page, Sergey Brin
Headquarters	Googleplex, Mountain View, California, U.S.[3]
Area served	Worldwide
Key people	Larry Page (CEO) Eric Schmidt (Chairman) Sergey Brin (Director of Google X and Special Projects)[4]

Send us your feedback.

Print Print this page

Permalink <http://dw.de/p/1ENVV>





Por que NLP é difícil?

- Ambiguidade
- Linguagem não-formal: gírias, neologismos
- Ironia
- Idiomas complexos:
 - Alemão: Donaudampfschiffahrtsgesellschaftskapitän (5 palavras)
 - Chinês: 50.000 caracteres diferentes (2-3 mil em um jornal)
 - Japonês: 3 sistemas de escrita
 - Etc

Write Traditional “If-Then-Else” Rules?

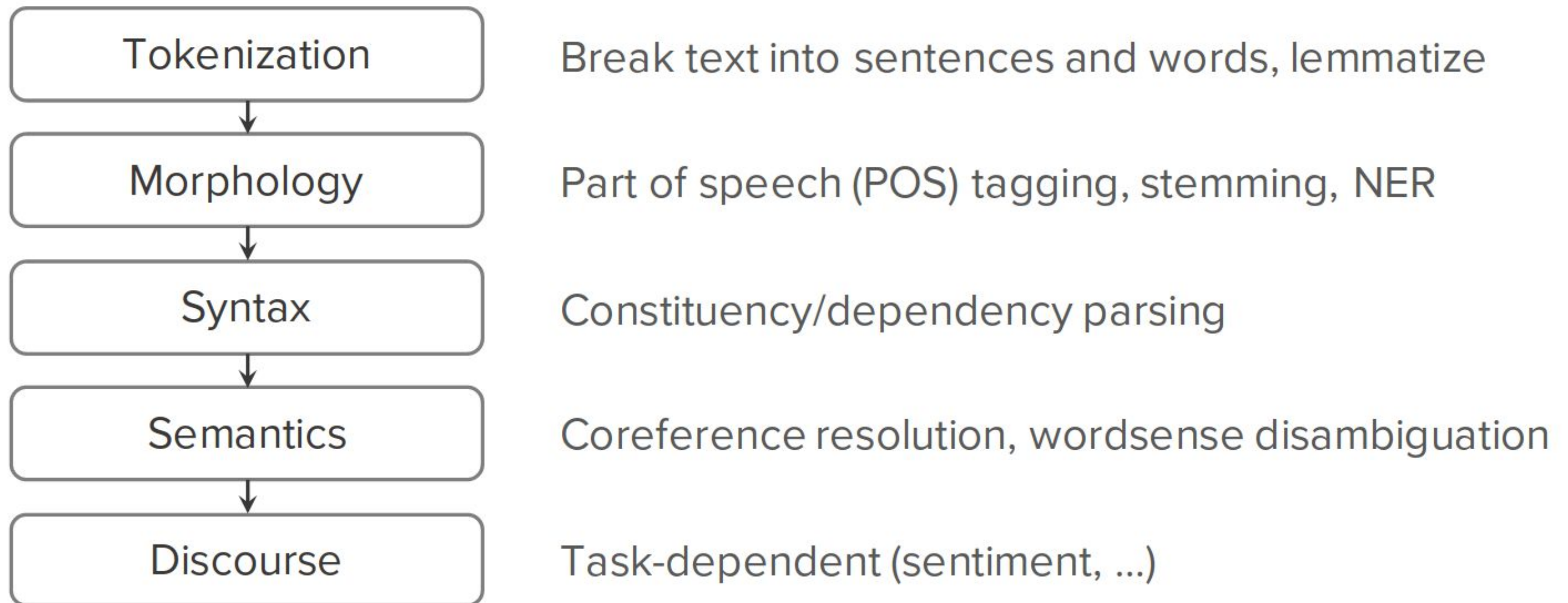
BIG NOPE!

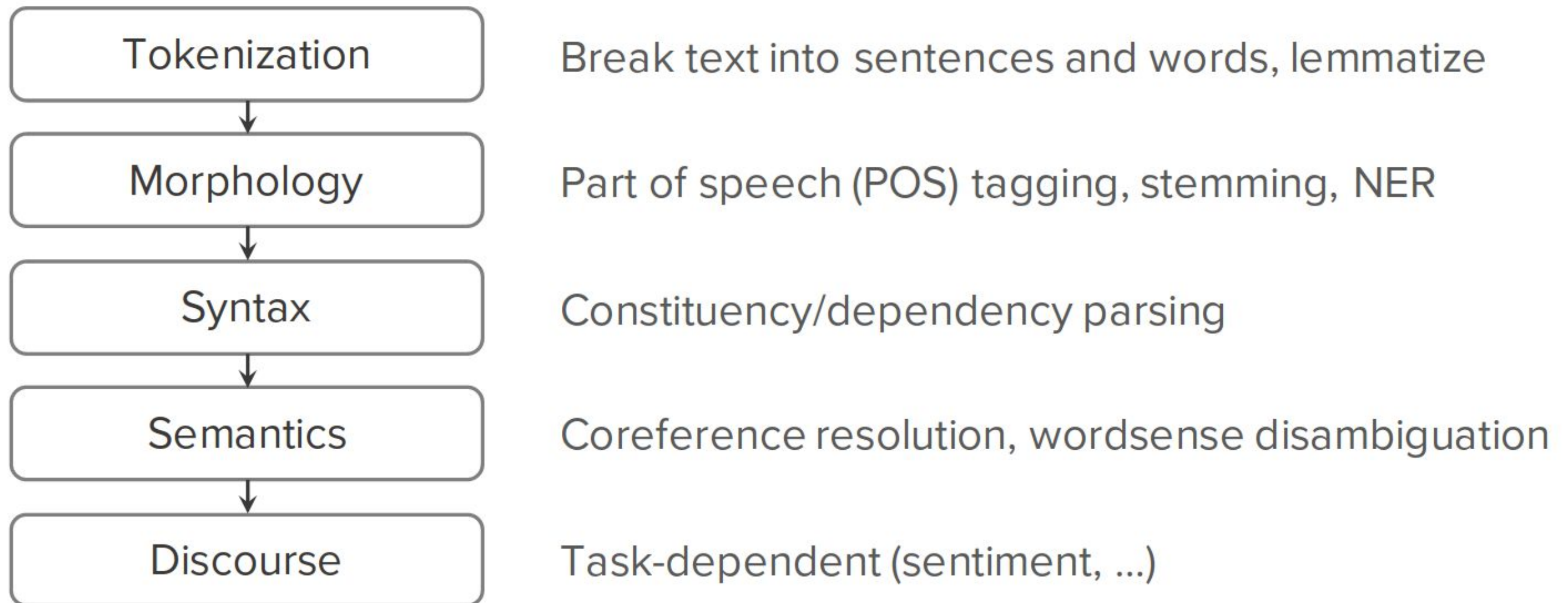
Leads to very large and complex codebases.
Still struggles to capture trivial cases (for a human).

Better Approach: Machine Learning

- “
- A computer program is said to learn from experience **E**
 - with respect to some class of tasks **T** and performance measure **P**,
 - if its performance at tasks in **T**, as measured by **P**,
 - improves with experience **E**.

— Tom M. Mitchell





Interessado?

- <https://github.com/laugustyniak/awesome-sentiment-analysis>
- https://github.com/Kyubyeong/nlp_tasks