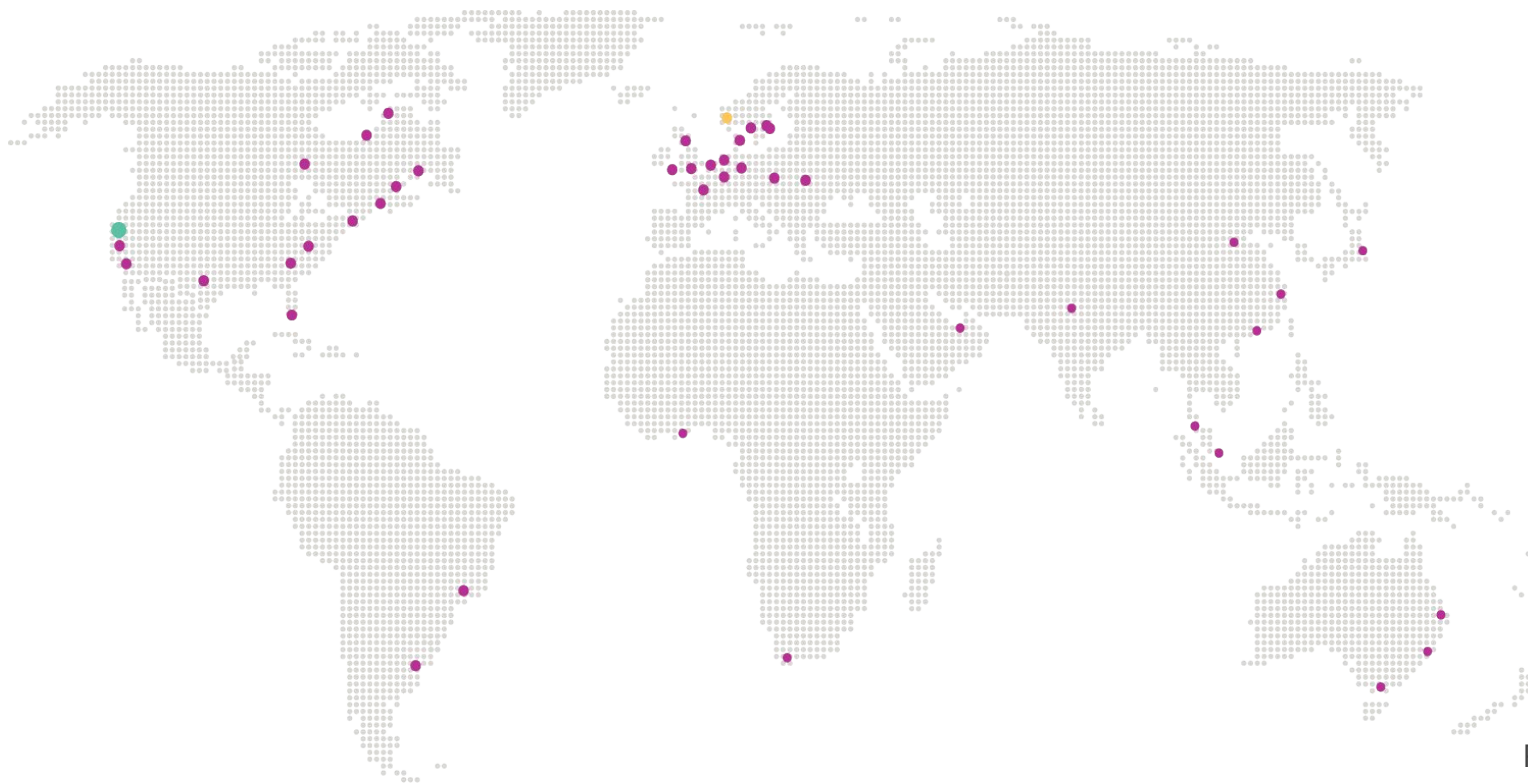# NLP Stack
At Meltwater

Miháltz Márton
marton.mihaltz@meltwater.com

Meltwater At a Glance

30,000+
Clients
in 121 countries

60+
Offices
In 27 countries

280+
Client Success Managers

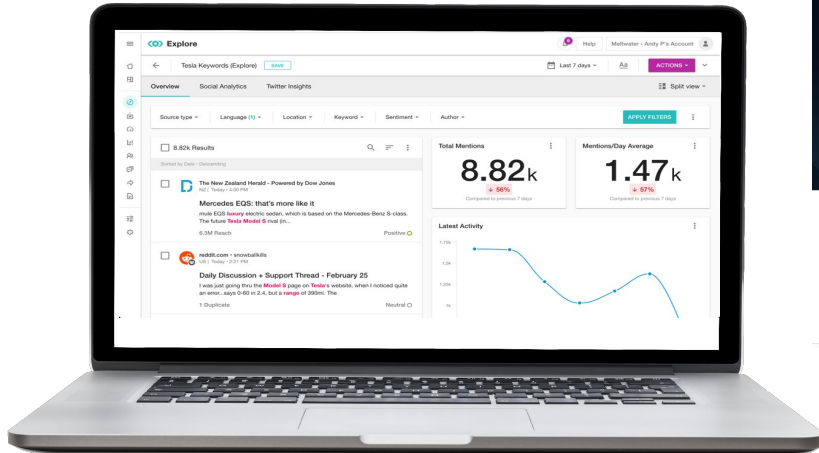**200+**
CS Support

180+
Product Developers

‹o› Meltwater

About Meltwater

- Media Intelligence SaS
  - Media monitoring
  - Audience insights
  - Competitive intelligence
  - Social media engagement
  - Influencer marketing

Launch
Apple

Apple has had a launch related event

MyBroadband Bloomberg
Oct 7 • 5:59 AM

Apples reveals launch date for 5G iPhones

Apple Inc. announced that its biggest product launch event of the year will be held Oct. 13. The Cupertino, California-based technology...

266k Reach | Neutral

Acquisition
Cisco | IMImobile

Cisco & IMImobile are involved in an acquisition event

CISION **PR Newswire**
Feb 19 • 1:30 PM

Cisco Completes Acquisition of IMImobile PLC

SAN JOSE, Calif., and LONDON, Feb. 19, 2021 /PRNewswire/ -- News Summary: Cisco has completed the acquisition of IMImobile PLC. ...

8.49M Reach | Neutral

Recognition
HCL | Cisco

Data Flow

- ## 10M sources globally
  - News, social networks, comments, blogs, forums, reviews
  - Print media, broadcasts, podcasts
  - 1B docs / day
- ## ES cluster
  - $1.4*10^{12}$ documents (2 years rolling)
- ## NLP Enrichments
  - 19 languages, 170+ services
  - 10B requests / day
- ## Knowledge Graph
  - Nodes: organizations, key persons, industries, stock indices, addresses etc.
  - Edges: relationships (affiliations, subsidiaries, industry associations etc.) or events (mergers and acquisitions etc.)

NLP Services

- Language Detection
- Analyzer (tokenizer, PoS, stemming, stopwords, sentence splitter, chunking)
- Charikar LSH
- Keyphrase Extraction
- Document Categorizer
- NER
- NED
- Sentiment Analysis
- Event Detection

Deep Learning Models

- TensorFlow
  - Sentiment: GloVe + CNN
  - NER: GloVe + BiLSTM
- Training, Hyperparameter Optimization: AWS SageMaker
- CPU inference!
  - 250K req./s, <10ms avg. resp. time 24/7
  - Save $$$
- Improving models
  - Customer sentiment overrides
  - Data augmentation and retraining
- Adding new languages
  - Luigi data pipeline: balanced collection estimated via En MT
- Models serving: Python -> Java
  - 4-5x fewer resources, 2-3x faster

Service Infrastructure

- Dockerized services
  - Python: Falcon, Gunicorn
  - Java: Spring Boot, Jetty
- Total 10B service calls / day
- AWS ECS
  - EC2 Spot instances
  - 2 clusters, 172 services, 396 tasks, ~1100 cores
- Benthos streaming pipeline + Kafka topics + Elastic Search
- Drone, Kibana, Prometheus, Grafana, Opsgenie
- Total Terraform(™)
- Moving to Kubernetes
  - MW's own k8s cluster on AWS

More information

Under The Hood Meltwater Engineering Blog

http://underthehood.meltwater.com/

Meltwater