Entity Extraction: Our Approach

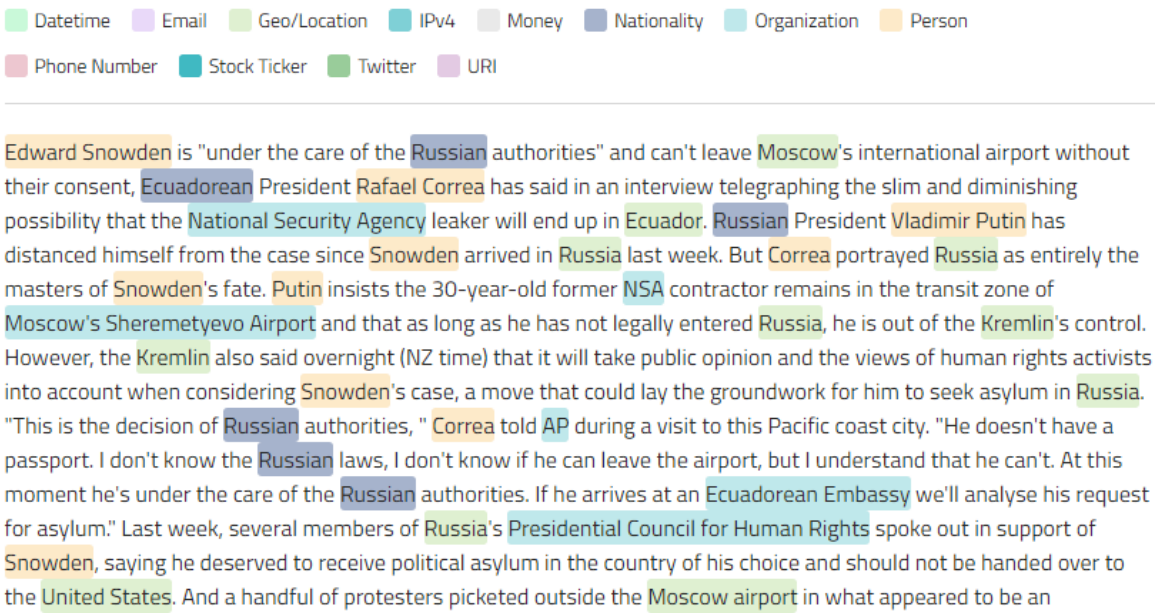
Finch for Text

At Finch Computing, we build new ways of interacting with information. Perhaps nowhere is this more apparent than in our text analytics solution, Finch for Text® which makes human-generated text machine-readable. We say Finch for Text® is “software that reads and reasons” because proprietary technologies in the product enable it to extract, disambiguate and enrich entities and to assign sentiment to these entities in ways other solutions just can’t replicate. Below is a sampling of how we approach entity extraction in particular in support of a number of business and mission-critical use cases.

What is entity extraction?

Entity Extraction is the process of extracting named entities – like people, places, organizations and more – that appear in structured or unstructured text documents. We use a combination of proprietary and licensed text analytics models to correctly isolate these entities in text, and to categorize them according to their type.

Below is a screen shot of our Finch for Text® demo. Out of the box, its models are trained on news content and enable the product to extract more than 20 types of entities and entity subclasses. For example: identifying a mention of the National Security Agency, below, as an entity whose type is: Organization, and whose subclass is: Government Agency.



Custom extractions are available in Finch for Text®. The product can be tuned for a particular domain and configured to support custom taxonomies, dictionaries and patterns. We have developed models for cyber security and finance for specific customer use cases, for example. Product extractions are coming soon, among other additions to our supported entity types.

Entity Extraction: Our Approach

What's Unique About Our Approach?

Finch for Text® leverages a patent portfolio rich in innovations in topic modeling, text analytics, pattern detection and more to enable it to quickly and accurately extract entities from within huge volumes of structured and unstructured text. This enables it to take a context-based approach, understanding not just the entities mentioned, but their surrounding context in order to appropriately identify them and categorize them for analysis.

Assessing Extraction Performance

In January 2017, we performed a series of competitive benchmarking tests to understand how Finch for Text® compares to 14 of the most popular text analytics solutions – products like NetOwl, AlchemyAPI (now part of IBM Watson) Lexalytics, and Microsoft and Google's beta offerings.

Finch for Text	PEOPLE	ORG	GEO
	90/90	91/76	84/89

Finch for Text won every head-to-head competition, across every entity type.

Above are our precision and recall results (expressed as P/R in percentages) across PEOPLE, ORGANIZATIONS and GEOGRAPHIC PLACES. Precision and recall are common metrics used to evaluate extraction quality. Precision = How much did the solution get right; Recall = Did the solution catch everything it was supposed to.

To perform our comparison, we tested Finch for Text® and the other 14 solutions on an identical, 400-document corpus of news and social media content – precisely because it varies in topics, entities, length and more. It's a perfect example of a streaming, human-generated content feed – not unlike the types of content enterprises need to understand every day – emails, research reports, message traffic, etc. And, again, Finch for Text® won every time.

How is entity extraction used in real-world scenarios?

Commercial and government organizations alike are using the extraction capabilities in Finch for Text® to support a variety of critical text analytics functions. These include:

- Tagging content from a real-time stream with no volume or capacity limitations.
- Organizing massive content libraries or archives.
- Developing ontologies to govern text classification and tagging projects.
- Improving the ability to find content from within huge, enterprise-volumes of text.