Entity Enrichment: Our Approach

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 enrichment in support of a number of business and mission critical use cases.

What is entity enrichment?

Entity enrichment is a complement to the text analytics functions of extraction and disambiguation. Entity extraction involves isolating an entity in text and determining its type (person, place, etc.); and disambiguation involves resolving the identity of a particular entity to a knowledgebase so that a user can differentiate between two identically named entities of the same type, such as Paris, Texas versus Paris, France.

Taking things a step further, entity enrichment involves adding additional data *about* an entity to it so that it is richer and more valuable for further analysis. For example, understanding that a business is also affiliated with a certain industry, has a certain URL or street address, uses a certain ticker symbol or social media handle, has a certain CEO and board members all make that entity more meaningful. From these enrichments, you can make connections in the data and link two seemingly disparate things to one another based on a characteristic they share – even if it's not expressly mentioned in the text.

Example:

From just this single entity: Amazon, Inc.

We can add all of this metadata, and more:

- Entity Type: Organization
- Class: Business
- Type: Online Retailer
- Ticker Symbol: (AMZN) NASDAQ
- URL: www.amazon.com
- IP Address: 184.31.198.208
- CEO: Jeff Bezos
- Date Founded: 7/6/1994
- Address: 1516 Second Ave. Seattle, WA 98101
- Latitude: 47.6227
- Longitude: -122.3362

Example:

From just this single entity: www.finchcomputing.com

We can add all of this metadata, and more:

- Entity Type: URL
- Host Name: finchcomputing.com
- IP Address: 52.206.117.180
- Admin 1 ISO: VA
- Admin 1 Label: Ashburn
- Continent: North America
- Country ISO: US
- Country Label: United States
- Latitude: 39.0481
- Longitude: -77.4728
- Metro Code: 511
- Postal Code: 20149
- Time Zone: America/New York

Example:

From just this single entity: Wrigley Field

We can add all of this metadata, and more:

- Entity Type: Geographic Place
- Type: Recreational Park
- City: Chicago
- County: Cook
- State: IL
- Country: USA
- Latitude: 41.9481
- Longitude: -87.6553

Entity Enrichment: Our Approach

What types of entities do we enrich?

Finch for Text® can enrich more than 20 entity types out of the box in real-time. Custom enrichments can be performed to satisfy a use case within a particular domain, such as cybersecurity (hacking tools and techniques, hacking groups, malware, bad actors, etc.) and banking (credit cards, IIN and BIN numbers, bank names, routing numbers, etc.)

We've curated huge knowledgebases of people, places and organizations to support this capability. And we continually update our licensed sources and publicly available sources of information to add to our enrichment offerings.

Out of the box, Finch for Text® can enrich multiple entity types:

- People
- Places
- Organizations
- IP Addresses
- Hostnames
- URLs
- Phone Numbers
- Ticker Symbols
- Monetary Values
- And more...

How does it work?

Finch for Text® uses natural language processing, machine learning, sophisticated statistical models and other heuristics to perform entity extraction and disambiguation. Once an entity's identity is resolved, we can cross reference it with our massive knowledgebases of people, places, organizations and other entities to add additional metadata about those entities. (We can do the same thing with our customers' knowledgebases, or add their entities to ours.) The enrichments are returned in JSON via API to customers who use the outputs within their existing analytics platforms or to build their own custom applications.

The sources from which we pull content to curate our knowledgebases and to enrich entities for our customers include:

- Wikipedia/Wikidata
- DBpedia
- Twitter
- Corporate & Gov't Homepages
- Paid Content Sources
- US Geological Survey (USGS)
- Nat'l Geospatial-Intelligence Agency (NGA)
- And more...

How can it be used?

Customers are using our entity enrichment capabilities to make connections in their unstructured data like never before. These include:

- Enriching streaming content feeds in real-time and at large scales
- Enriching content in a data lake or Hadoop cluster to prep it for meaningful analysis
- Linking content in massive content libraries so that it is more easily searchable and accessible
- Monitoring message traffic on the dark web to understand criminal networks and prevent fraud
- Improving research and intelligence gathering efforts by making contextual information about an entity instantly available to an analyst

To learn more, please contact us at sales@finchcomputing.com or visit us at www.finchcomputing.com.