Entity Disambiguation:
Better Search, Better Results

Entity disambiguation refers to the ability to resolve an entity's identity to a knowledgebase. It is not merely entity-type classifying – as in determining that a reference to "George Washington" is a reference to the person and not the bridge in New York City, for example. Instead, entity disambiguation involves correctly distinguishing between two identically named entities of the same type – as in, John Roberts the Chief Justice of the U.S. Supreme Court, John Roberts the Fox News correspondent, or any one of the hundreds of individuals in the world named John Roberts.

Doing this requires understanding entities and their surrounding context. It requires understanding nuance and inferences, nicknames and short-hands, misspellings and variances in upper and lower case text. It also involves the development and curation of massive knowledge bases, full of entities - *and knowledge about those entities* - in order to correctly interpret their surrounding context in a piece of text.

Many Natural Language Processing solutions will claim to do entity disambiguation but either use a modified definition of the practice, or concentrate only on a particular domain or entity type; even worse, they deliver accuracy that is sub-par at best. Our approach is more precise and, as such, it improves search. It reduces the number of missed entities in searches and the number of duplicates, as an example, we resolve multiple types of mentions to the same entity, as an example: Donald Trump, Trump, President Trump, Donald J. Trump, etc.

Our Proprietary Approach Drives Accuracy, Removes Noise:

Finch for Text® takes a proprietary, context-based approach to entity disambiguation. We hold patents that involve improving traditional, algorithmic approaches in order to assess more components of a document. As such, Finch for Text® can understand entities, key phrases, topics and embedding vectors in order to link an entity appearing in text to our knowledgebase. This essentially creates a mathematical feature model for every piece of text and allows us to quickly and intimately understand the entities referenced in it.

Our patented approach works across any language. We turn documents into feature models in order to understand entities and their associations in entirely new ways. Resulting in accurate extraction and disambiguation of people, organizations and geographic places.

Our Knowledgebases

At Finch Computing, we have developed huge knowledgebases of people, places and organizations. Our knowledgebase is always growing and includes hundreds of millions of discrete entities. However, what's of the most value is the knowledge we have *about* those entities – knowledge of the topics associated with an entity, facts about an entity, key phrases that often appear with or near an entity.

Using the John Roberts example, our knowledgebase entry for the Chief Justice would include terms like law, judge, legal, courts, rulings, etc. These terms comprise a numeric, machine-readable "topic" that is linked to the John Roberts entity in our knowledgebase. Key phrases like "the nine justices" or "in a historic ruling" are similarly linked to the entity. Facts about Mr. Roberts – like where he went to school, when he was born, past positions he's held – are also stored in our knowledgebases.

We use this information to derive features about an entity, and all of this information – billions of pieces of knowledge – are immediately at our disposal when analyzing a piece of unstructured text. We constantly add to our knowledgebases with human and machine driven updates. And we can add customer-specific information to it with ease.

See below for an example of how Finch for Text uses context and mathematical feature models to correctly distinguish one "John Roberts" from another.

ENTITY TYPE	# OF ENTITIES
People	1.9 M
Geographic Places	40M
Organizations	2.1M

Companies (All NYSE, NASDAQ & Russell 3000)
Educational Organizations
Government Organizations
News Outlets
Performing Organizations
Criminal Organizations
Social Media Companies
Sports Organizations
Trust/Funds

Attributes about entities help drive accuracy and fidelity. We have millions in our knowledgebase.

ATTRIBUTES	# IN OUR KB
Stock Tickers	75,231
Twitter Handles	455,205
PermIDs	1,228,806
LEIs	169,104
CIKs	76,020
Crunchbase	579,024
LinkedIn	268,511
Freebase	1,716,137
Facebook	348,482
Wikidata	1,793,472

John Roberts

John Glover Roberts Jr. (born January 27, 1955) is an American lawyer and jurist who serves as the chief justice of the United States. Roberts has authored the majority opinion in several landmark cases, including Shelby County v. Holder, National Federation of Independent Business v. Sebelius, King v. Burwell, Department of Commerce v. New York, and Department of Homeland Security v. Regents of the University of California. He has been described as having a conservative judicial philosophy but has shown a willingness to work with the Supreme Court's liberal bloc, and since the retirement of Anthony Kennedy in 2018 has come to be regarded as a key swing vote on the Court. Roberts presided over the first impeachment trial of Donald Trump in early 2020; however, he declined to preside over the second impeachment trial of Trump, who was impeached as president, but whose term had expired by the time of the trial.

Extraction Confidence: 0.7961   Finch Id: 30416633857
Disambiguation Score: 0.5209   Perm Id: 34419704213
Salience Score: 0.8364

John Roberts

John David Roberts (born November 15, 1956) is a Canadian-American television journalist currently working for the Fox News Channel, as the co-anchor of "America Reports". Roberts formerly worked as the Fox News Chief White House Correspondent from 2017 to 2021, covering the Donald Trump presidency.

Extraction Confidence: 1.0000   Finch Id: 5483200513
Disambiguation Score: 0.7518   Salience Score: 0.7389

To learn more, contact: sales@finchcomputing.com