

Linking and learning for real-world data

Rosette® Entity Resolver (RES) reveals meaningful information in your text. It connects the words that represent real-world things to one another and to entities in an entity database like Wikipedia, both within and across documents.

Good quality entity resolution means dealing with three key problems: *variety*, where one thing can have many names; *ambiguity*, where many things can have very similar if not exactly the same name, and *ghosts*, where some collection of names identify a previously unknown real-world thing.

RES enriches your text with high, quality metadata, enabling you to perform intuitive, entity-centric search and discovery. With it you can power notification applications designed to detect and track new people in text streams. It provides excellent raw material for building the custom *knowledge graphs* at the heart of many of today’s most innovative applications.

KEY FEATURES

- Standard training from 2.5M Wikipedia entities
- “Learning” Mode: Identifies previously unknown entities (“ghosts”) and learns new aliases from text as it processes
- “Linking” Mode: Rapidly links only known entities
- Custom entity database training
- Fast and scalable
- Industrial-strength support
- Flexible and customizable
- Java
- Unix, Linux, Mac, or Windows

EXAMPLES:

Tamerlan Tsarnaev

Tamerlan Tsarnaev (TheAtlantic.com)

—or—

Tamerlane Tsarnaevy (Mir24.net)

Apple

Apple Corps Ltd. (Music)

—or—

Apple Inc. (Technology)

Paris

Paris, Texas (33°39 N, 95°32 W)

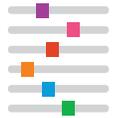
—or—

Paris, France (48°51 N 2°21 E)

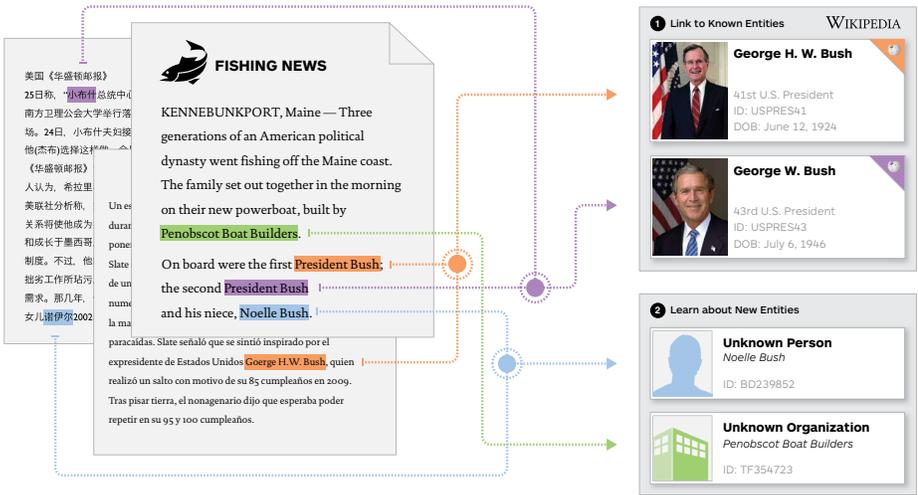
Start using RES

Try our free product evaluation

www.basistech.com



Organize Big Text using entity linking and learning



Measuring Confidence

Confidence measures are essential for effective use of statistically based systems. RES can be configured to deliver confidence measures with each of its clustering and linking decisions, allowing developers to use the RES output intelligently.



Rosette Entity Resolver can be run in one of two modes:

Linking

In linking mode, RES will link the names of people, places and organizations in the text to entities in the entity database.

Anything that can't be associated with an existing entity will be ignored.

This mode is optimized for high scale and stable throughput.

Learning

In learning mode, RES not only links names to known entities, but also discovers new entities mentioned in the text (often called "ghosts"), and remembers the new aliases and contexts it has found for all entities.

For example, once "J. Doe" has been encountered and linked to the "John Doe" entity, future occurrences of "J. Doe" will be matched with greater confidence.

RES comes pre-trained to link to a Wikipedia-derived 2M+ entity database. RES may be further trained by adding to this entity database or by providing an entirely new entity database.

What is training?

Training currently involves adding information about real-world entities to the system such as names, aliases, related entities, and example documents.

A simple example is adding a new alias to a Wikipedia-derived entity to improve resolution accuracy.

Under the hood

RES uses a machine-learned model to associate names and their contexts with collections of information drawn from the entity databases with known entities.

In linking mode, RES fixes both the number of entities and the information within.

Learning mode allows new entities to be created and new information to be added to existing entities. As this system state grows, RES intelligently prunes the information to maintain performance.

EXAMPLE:

Basketball player Jeremy Lin is often referred to as "Linsanity".

Training allows developers to add the "Linsanity" alias to the entry for Jeremy Lin. The next time "Linsanity" is encountered, it will be resolved appropriately.

Requirements

Code Base



Platform Support



Full Support Languages

- Arabic
- English
- Spanish
- Chinese

Partial Support Languages

- Russian
- Japanese
- Korean
- Pashto
- Persian (Dari)
- Persian (Farsi)
- Urdu