

ENTITY RESOLUTION

Principle-based entity resolution explained

How Senzing® software eliminates pre-training, tuning, and expert configuration — at any scale.

DOCUMENT TYPE

Technical whitepaper

APPLIES TO

Senzing SDK

AUDIENCE

Technical evaluators & architects

PUBLISHED

2019

Senzing® software uses a unique principle-based approach to entity resolution that eliminates the need for pre-training, tuning or experts. This general-purpose method can be used across a nearly unlimited range of entity types, such as people, organizations, vessels and vehicles.

Principles are just one of the reasons our software is so easy to deploy and use, while delivering higher quality and more accurate results than other entity resolution methods.

This document provides details about how principle-based entity resolution works. It explains the concept of using principles to perform entity resolution, how principles differ from traditional rules, why principles make the entity resolution process more efficient, and how they deliver more precise and accurate results.

BACKGROUND

What is principle-based entity resolution and how is it better?

Principles are a special form of generalized knowledge that draws on common truths or assumptions. The differences between the rules used by some entity resolution methods and the principles used in the Senzing software are distinct. Here's an example:

You tell your child to quit throwing rocks at cars — a rule. The next day you find him throwing baseballs at SUVs. Instead of adding more rules, why not one simple principle: don't throw things at other people's stuff?

In the Senzing software, principles are based on the expected behaviors of entity attributes, e.g., names, addresses and identifiers. For example, social security numbers (SSNs) typically point to only one person, but dates of birth (DOB) behave differently, as many people share the same DOB. There are always exceptions.

These exceptions are learned in real time, as new data is received. For example, when multiple people are using the same SSN, our software detects it, labels that SSN as generic and reevaluates all prior records with that number.

Our principle-based method assigns three behaviors to each entity attribute:



Frequency — does one, few, many or very many entities generally share the same value? **Exclusivity** — does an entity typically have just one such value? **Stability** — is this an exclusive value that is generally constant over an entity's lifetime?

The software comes preconfigured with the common attributes and expected behaviors of people and organizations. See Table 1 for some examples. You can start loading and resolving entities without any configuration, training or tuning. If you need to add a new attribute, such as an additional identifier, just add the name of the attribute and assign its three behaviors.

TABLE 1: EXAMPLES OF PRECONFIGURED ATTRIBUTES

ATTRIBUTES	FREQUENCY	EXCLUSIVITY	STABILITY
NAME	NAME	No	No
DATE_OF_BIRTH	FM	Yes	Yes
GENDER	FVM	Yes	No
ADDRESS	FF	No	No
PHONE	FF	No	No
SSN	F1	Yes	Yes
DRIVERS_LICENSE	F1	Yes	No
PASSPORT	F1	Yes	No
TAX_ID	F1	Yes	Yes
FACEBOOK	F1	No	No
EMAIL	F1	No	No
WEBSITE	FF	No	No
ACCOUNT_NUMBER	F1	No	No

CORE CONCEPT

How are entity resolution principles used?

Principles are used to determine when entities are the same, possibly the same or possibly related. Each principle is expressed as a combination of frequency, exclusivity and stability behaviors plus special processing for names.

Principles are obviously rules, but they are much more generalized than the type of rules typically used for record matching and linking. Senzing software uses about 35 general-purpose principles to resolve people, organizations, vessels and much more. Each principle contains logic to confirm or deny its use, resulting in either the principle being applied or not.

When an incoming record is received, the software quickly identifies potential candidate matches. The incoming record’s attributes are scored against the attributes of each candidate. These scores are then passed into the entity resolution principles, which determine the outcome for each candidate. See Table 2 below for some examples of the default principles included in the Senzing software.

Most users rely exclusively on our proven default principles. In rare instances, users may need to adjust principles to account for specific requirements. These changes can be easily added.

REFERENCE

Entity attribute behavior details

1. Frequency (of the entity attribute value)?

- a. **One** (or F1 for a frequency of one) – only one entity should have this value, usually an ID number, account number or email
 - b. **Few** (or FF for a frequency of few) – only a few entities should have this value, usually a method for contacting an entity, such as an address or home phone
 - c. **Many** (or FM for a frequency of many) – many entities can have this value, including attributes such as DOB, etc.
 - d. **Very many** (FVM for a frequency of very many) – an extremely large number of entities have this value, including attributes such as gender, citizenship, nationality, etc.
 - e. **None** – for attributes used internally that should not impact entity resolution
 - f. **Name** – a special frequency used for attributes containing an entity's name, such as person name or organization name
-

2. Exclusivity (entities generally only have one of these attributes)?

- a. **Yes** – for attributes that, if different, would strongly encourage the records to be kept apart, such as government-issued IDs, DOB, gender, etc.
 - b. **No** – for everything else
-

3. Stability (or extremely strong exclusivity throughout an entity's entire life)?

- a. **Yes** – for attributes that are exclusive, generally reliable, and extremely useful for differentiating between entities, such as DOB and SSN
 - b. **No** – for everything else
-

RESOLUTION LOGIC

Default principles in Senzing software

TABLE 2: EXAMPLES OF DEFAULT PRINCIPLES INCLUDED IN SENZING SOFTWARE

PRINCIPLE	SKIP ON CONFLICTING EXCLUSIVES*	DESCRIPTION	EXAMPLE
MATCHES			
SAME_F1+CLOSE_NAME	Yes	Two records sharing the same F1 and a close NAME will match unless they have conflicting exclusive attributes.	Two records for Bob Smith with the same account number match, but if they had different DOBs they would not. With different DOBs, they are more likely father and son.
MULT_FF+CLOSE_NAME	Yes	Two records sharing multiple FFs and a close NAME will match unless they have conflicting exclusive attributes.	Two records for Pat Smith with the same address and phone number match, but if they had different gender attributes they would not. With different genders, they are more likely Patricia and Patrick Smith.
MULT_STAB+LIKELY_NAME+SAME_F1	No	Two records sharing multiple stable attributes, a likely NAME, and the same F1 will match.	Two records for Pat Smith and Patricia Jones with the same DOB and the same SSN match, even if the passport is different. This is more likely to be someone who married and changed her last name.
POSSIBLE MATCHES			
CLOSE_NAME+SAME_F1	No	Two records sharing the same F1 and a close NAME will be possible matches even if there are conflicting exclusives.	Two records for Bob Smith with the same account number but different DOBs are possible matches.
CLOSE_NAME+CLOSE_STAB	No	Two records sharing a close stable and close NAME will be possible matches.	Two records for Bob Smith and Robert E Smith with DOBs of 4/5/1963 and 4/5/1968 are possible matches
POSSIBLY RELATED			
SAME_F1	No	Two records sharing the same F1 but different NAMES and other information.	Two records for Bob Smith and Jack Williams with the same account number are possibly related.
CFF_SURNAME	No	Two records sharing a close FF and family NAME will be possible matches.	Two records for Bob Smith and Fred Smith with the same address are possibly related.

**Specifies whether a principle should be skipped when its qualifications are met but an attribute exhibits conflicting exclusive behavior, e.g., different dates of birth.*

VERSATILITY

Principles apply across entity types

Many principles can be applied to other entity types besides people. Below are examples of how some of the principles in Table 2 can be used to resolve organizations and vessels.

Organizations

- Match – two records for ABC Company with the same email address
- Possible Match – two records for ABC Company and ABC Enterprises Inc. with the same address and different tax IDs
- Possibly Related – two records for ABC Company and XYZ Enterprises Inc. with the same address and phone number

Vessels

- Match – two records for the vessel Maltego with the same Maritime Mobile Service Identity (MMSI) number
- Possible Match – two records for Maltego with the same MMSI but different vessel types
- Possibly Related – two records for vessels named Maltego and Seven Seasons with the same MMSI number

CONTINUOUS IMPROVEMENT

Senzing principles and accuracy improve over time

The Senzing team is continuously exploring new ways to improve our principles, but we rarely find the need to change them. When we do change a principle to improve the results of a certain entity type, we usually find that other entity types also benefit. For example, a change that improves accuracy when resolving people typically improves accuracy when resolving cars, boats and planes too.



For more information about all of the unique capabilities in Senzing software, read our [white paper](#) or visit the [technology page](#) on our website. To get started right away, get a [free download](#).

We hope you enjoy our technology.

1. <https://senzing.com/uniquely-senzing>
2. <https://senzing.com/our-technology>
3. <https://senzing.com>