



# Rosette Name Indexer

## Comparison to Common Alternatives

The Name Indexer component of Rosette® by Babel Street was evaluated in December 2019 against three common alternatives using a dataset with 7,571 names, with at least 10 variants for each name.

These alternatives included:

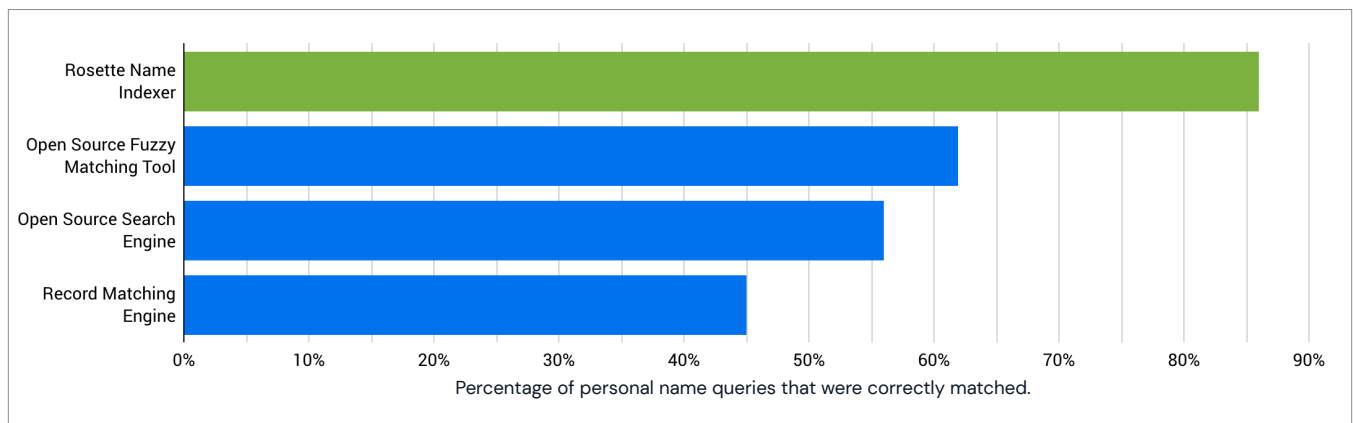
- An open source fuzzy matching tool
- An open source search engine
- A record matching engine

Testing and analysis show that these alternatives fall short of Rosette because they lack script/language

support, lack essential name phenomena support, and use rigid or overly simplified methods to calculate match scores.

### Superior accuracy

Where a correct match is defined as matching a “gold standard” version of a name to one of its variants, Rosette outperforms the alternatives by 24% or more for person name matching.



## Coverage of match phenomena

	OS Fuzzy Matching	OS Search Engine	Record Engine	Rosette
<b>Exact Match</b> (two names are identical “John Smith” ↔ “John Smith”)	✓	✓	✓	✓
<b>Normalization</b> (ability to identify matching names whose characters normalize to the same letters “LINDSTROM-JONES” ↔ “Lindström-Jones”)	Partial	✓	Partial	✓
<b>Stop Words</b> (ability to remove “noise words” from names “Dr.” ↔ Mr. ↔ Ph.D.”)	None	None	None	✓
<b>Nicknames</b> (ability to recognize common nicknames such as “William ↔ Will ↔ Bill ↔ Billy”)	None	None	Partial	✓
<b>Fuzzy Match</b> (statistical model for fuzzy matching)	None	None	None	✓

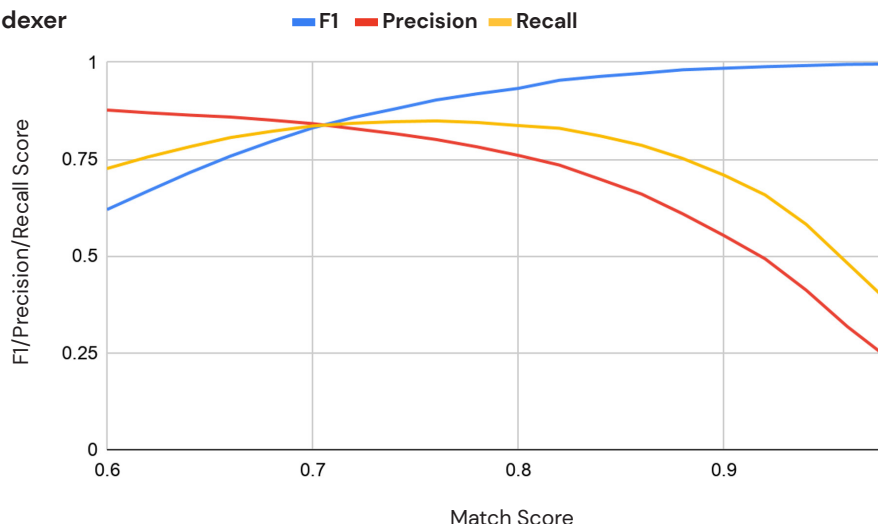
	OS Fuzzy Matching	OS Search Engine	Record Engine	Rosette
<b>Truncation</b> (ability to recognize long names cut short "Blankenship" ↔ "Blankensh")	Partial	✓	✓	✓
<b>Cross-lingual</b> (ability to match the same name written in different languages and scripts "Mao Zedong" ↔ "Мао Цзэдуи" ↔ 毛泽东 ↔ 毛澤東)	None	None	Partial	✓
<b>String Similarity</b> (ability to detect similarity due to edit distance "John" ↔ "Jhon")	None	✓	None	✓
<b>Missing Name Components</b> (ability to take into consideration a missing name component "Phillip Charles Carr" ↔ "Phillip Carr")	✓	✓	Partial	✓
<b>Out-of-Order Deletion</b> (ability to take into consideration a missing name component in conjunction with other name components having moved "Phillip Charles Carr" ↔ "Carr Charles")	✓	✓	Partial	✓
<b>Organizational Aliases</b> (ability to handle organizational name acronyms "Very Fine Groceries, LLC" ↔ "VFG")	None	None	None	✓
<b>Initials</b> (ability to handle replacement of a name with an initial "J.E. Smith" ↔ "James Earl Smith")	Partial	✓	Partial	✓
<b>Reordering</b> (ability to consider components that are a match, but penalize for a mismatch in the order of components "Diaz, Carlos Alfonso" ↔ "Carlos Alfonso Diaz")	✓	✓	✓	✓
<b>Insert Spaces</b> (ability to handle name components that appear to have been "glued" together "MaryEllen" ↔ "Mary Ellen" ↔ "Mary-Ellen")	Partial	Partial	None	✓
<b>Rotation</b> (ability to avoid over-penalizing for reordered name components "Carlos Alfonso Diaz" ↔ "Alfonso Carlos Diaz")	✓	✓	✓	✓
<b>Concatenation</b> (ability to consider if concatenating tokens produces a better match "Fred Will Sun" ↔ "Fred Wilson")	None	Partial	Partial	✓
<b>Gender Mismatch</b> (ability to detect when a male name is being compared to a female name and adjust the score accordingly "Jon Smith" ↔ "John Smith" but not "Joan Smith")	None	None	None	✓

## Useful match scores

Rosette outputs a nuanced match score as a decimal between 0 (no match) and 1 (perfect match). This match score can be used to balance precision and recall when setting a threshold score — the score above which names are considered "a match."

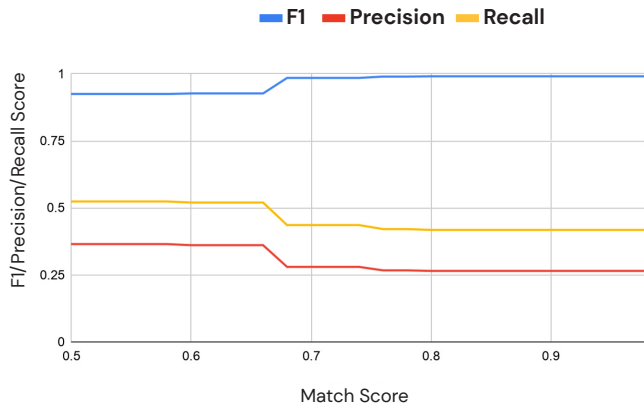
As the graph below shows, the precision and recall of Rosette meet at a point around a match score of 0.72. Users of Rosette can set a low threshold score to see more possible matches, and a higher score for only the most similar matches.

Rosette Name Indexer

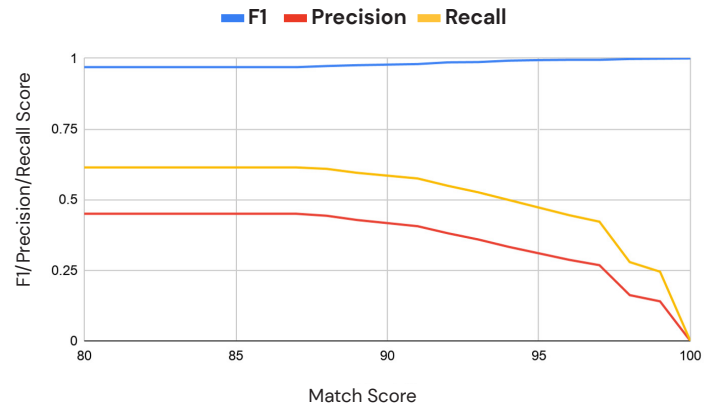


By contrast, the open source matching tool and the record matching engine operate in a binary “match” (score=1) or “no match” (score=0) paradigm without a range to indicate degrees of match. In this case, it is less clear what threshold will produce the desired balance of precision versus recall.

### Open Source Fuzzy Matching Tool



### Record Matching Engine



Furthermore, the open source search engine does not provide a comparable score. Thus it is not possible to compare match scores across multiple queries or configure business logic around the results.

*All names, companies, and incidents portrayed in this document are fictitious. No identification with actual persons (living or deceased), places, companies, and products are intended or should be inferred.*

Babel Street is the trusted technology partner for the world’s most advanced identity intelligence and risk operations. The Babel Street Insights platform delivers advanced AI and data analytics solutions to close the Risk–Confidence Gap.

Babel Street provides unmatched, analysis-ready data regardless of language, proactive risk identification, 360-degree insights, high-speed automation, and seamless integration into existing systems. We empower government and commercial organizations to transform high-stakes identity and risk operations into a strategic advantage.

**Learn more at [babelstreet.com](https://babelstreet.com).**