

International Journal of Population Data Science

Journal Website: www.ijpds.org



UK Address Matcher: High performance free software that uses ensemble techniques to maximise accuracy

Robin Linacre¹ and Thomas Hepworth¹

¹UK Ministry of Justice, London, United Kingdom

Existing address matching solutions employ widely differing methodologies, suggesting no single technique universally outperforms others and indicating that an ensemble approach may yield optimal results. High-performance parallelised engines such as DuckDB now enable simpler, faster implementations, making ensemble approaches more feasible. We aimed to develop free software prioritising ease of setup and speed, capable of matching against the full UK address list (over 30 million records) on standard hardware without complex database infrastructure, while achieving accuracy competitive with open source alternatives. We reviewed existing solutions to identify compatible methodological approaches, identifying several complementary techniques: probabilistic matching based on the Fellegi-Sunter model (Linacre et al., 2022) as implemented in Splink; signature-based matching using probabilistic signature methods (Zhang et al., 2017) as implemented in Whereabouts; reverse indexing similar to the ONS

Address Index Matching Service; and additional novel techniques. These methods were integrated into a unified ensemble framework, implemented as an open source Python library called `uk_address_matcher`, available through PyPI. Performance testing demonstrates that `uk_address_matcher` accuracy is competitive with existing open source tools. The software matches thousands of addresses per second on standard laptop hardware, while being easier and cheaper to set up. Conclusions and Implications Address matching is a common problem yet notoriously difficult due to inherent variability in address data. Successful linkage enables valuable applications from operational cost savings through deduplication to geospatial enrichment enabling spatial analysis. The `uk_address_matcher` package aims to democratise high-performance address matching by lowering technical barriers, enabling broader adoption without specialised infrastructure.

