Approximate String Matching

Luiz Celso Gomes Jr

Laboratory of Information Systems
Instituto de Computação, UNICAMP

Sep, 2013

Matching

- Microsotf == Microsoft
- Microsoft Windows == Windows (Microsoft)
- Microsoft Corporation == Micro Corporation || Microsoft Corp.
- sight == cite
- MS == Microsoft

Matching Functions

- Microsotf == Microsoft → Edit Distance
- Microsoft Windows == Windows (Microsoft) → n-gram blocks/Jaccard similarity
- Microsoft Corporation == Micro Corporation || Microsoft Corp. → block weighting/Weighted Jaccard similarity
- sight == cite → soundex
- MS == Microsoft → dictionary expansion