# Information Retrieval Spelling Correction Lecture-12

**Prepared By**

Dr. Rasmita Dash & Dr. Rasmita Rautray

Associate Professor

Dept. of CSE

# Content

- Forms of spelling corrections
- Isolated-term correction
- Edit distance
- Levenshtein Edit distance

# Forms of spelling corrections

We focus on two specific forms of spelling correction that

- *isolated-term* correction

- *context-sensitive* correction

In isolated-term correction, we attempt to correct a single query term at a time even when we have a multiple-term query.

**The carot example demonstrates this type of correction.**

Such isolated-term correction would fail to detect, for instance, that

**"The query flew form Heathrow"**

contains a mis-spelling of the term from - because each term in the query is correctly spelled in isolation.

# Types of isolated-term correction

- Edit distance
- K-gram overlap.

# Edit Distance

- The edit distance between string s1 and string s2 is the
- minimum number of basic operations that convert s1 to s2.
- Levenshtein distance: The admissible basic operations are
- insert, delete, and replace
- Levenshtein distance dog-do: 1
- Levenshtein distance cat-cart: 1
- Levenshtein distance cat-cut: 1
- Levenshtein distance cat-act: 2

# Levenshtein distance: Algorithm

$\textsc{Levenshtein Distance}(s_1, s_2)$

1   **for** $i \leftarrow 0$ **to** $|s_1|$
2   **do** $m[i, 0] = i$
3   **for** $j \leftarrow 0$ **to** $|s_2|$
4   **do** $m[0, j] = j$
5   **for** $i \leftarrow 1$ **to** $|s_1|$
6   **do for** $j \leftarrow 1$ **to** $|s_2|$
7       **do if** $s_1[i] = s_2[j]$
8           **then** $m[i, j] = \min\{m[i\text{-}1, j]+1, m[i, j\text{-}1]+1, m[i\text{-}1, j\text{-}1]\}$
9           **else** $m[i, j] = \min\{m[i\text{-}1, j]+1, m[i, j\text{-}1]+1, m[i\text{-}1, j\text{-}1]+1\}$
10  **return** $m[|s_1|, |s_2|]$

Operations: insert (cost 1), delete (cost 1), replace (cost 1), copy (cost 0)

# Each cell of Levenshtein matrix

| cost of getting here from my upper left neighbor (copy or replace) | cost of getting here from my upper neighbor (delete) |
| --- | --- |
| cost of getting here from my left neighbor (insert) | the minimum of the three possible "movements"; the cheapest way of getting here |

# Levenshtein distance: Example

| | | f | | a | | s | | t | |
|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 |
| c | 1 / 1 | *1* / 2 | *2* / *1* | 2 / 2 | 3 / 2 | 3 / 3 | 4 / 3 | 4 / 4 | 5 / 4 |
| a | 2 / 2 | 2 / 3 | 2 / 2 | *1* / *3* | *3* / *1* | 3 / 2 | 4 / 2 | 4 / 3 | 5 / 3 |
| t | 3 / 3 | 3 / 4 | 3 / 3 | 3 / 4 | 2 / 2 | *2* / *3* | *3* / *2* | *2* / *3* | 4 / 2 |
| s | 4 / 4 | 4 / 5 | 4 / 4 | 4 / 5 | 3 / 3 | 2 / 4 | 3 / 2 | *3* / *3* | *3* / *3* |

Thank You