

# SPELL CORRECTION

WE WANT TO CORRECT THE DOCUMENTS  
BEING INDEXED AND TO CORRECT THE  
QUERIES (IN ORDER TO RETRIEVE THE  
CORRECT ANSWER)

## TWO APPROACHES

### ISOLATED WORD CORRECTION

check each word  
on its own for  
misspelling

### CONTEXT SENSITIVE CORRECTION

check misspelling /  
based on the  
surrounding words

## ISOLATED WORD CORRECTION

Given a lexicon containing correct spellings and  
a query (single word), RETURN THE WORDS IN THE LEXICON  
CLOSEST TO THE QUERY

SEVERAL  
MEASUREMENTS  
OF  
DISTANCE

EDIT DISTANCE

WEIGHTED E. D.

N-GRAM INDEX

NOTE

Not very effective, because some words may seem misspelled, but they are not

("from" or "form"? DEPENDS ON THE CONTEXTS)

### CONTEXT SENSITIVE CORRECTION

It is easy, from the context, to detect the misspell in the following phrase

"I flew form Pisa"  
from<sup>to</sup>

This is usually done by using heuristics such as

→ Return alternative popular docs

→ Query log analysis & tweaking

(check if it is similar to a popular query)