

Information Retrieval

Topic-Wild Card Query Processing

Lecture-10

Prepared By

Dr. Rasmita Dash & Dr. Rasmita Rautray

Associate Professor

Dept. of CSE

Content

- Introduction
- Types of wildcard queries
- How to process wildcard queries
- Permuterm Index

Wildcard queries

Wildcard queries are used in any of the following situations:

- the user is uncertain of the spelling of a query term
(e.g. Sydney vs. Sidney,
which leads to the wildcard query S*dney)
- the user is aware of multiple variants of spelling a term and
(consciously) seeks documents containing any of the variants
(e.g., color vs. colour)
- the user seeks documents containing variants of a term that
would be caught by stemming, but is unsure whether the
search engine performs stemming
- (e.g. judicial vs. judiciary,
leading to the wildcard query judicia*)
- the user is uncertain of the correct rendition of a foreign word
or phrase (e.g., the query Universit* Stuttgart).

Types of wildcard queries

- Trailing wildcard queries -- mon*
- Leading wildcard queries -- * mon
- General wildcard queries -- m * n

How to process wildcard queries

mon^* : find all docs containing any term beginning with *mon*

- Easy with B-tree dictionary: retrieve all terms t in the range:
 $mon \leq t < moo$

$*mon$: find all docs containing any term ending with *mon*

- Maintain an additional tree for terms backwards
- Then retrieve all terms t in the range:
 $nom \leq t < non$

Result:

- A set of terms that are matches for wildcard query
- Then retrieve documents that contain any of these terms

How to handle * in the middle?

- Example: m*nchen
- We could look up m* and *nchen in the B-tree and intersect the two term sets.
- Expensive
- Alternative: **permuterm** index
- Basic idea: Rotate every wildcard query, so that the * occurs at the end.

Permuterm Index

- For term hello: add
- *hello\$*

ello\$h

llo\$he

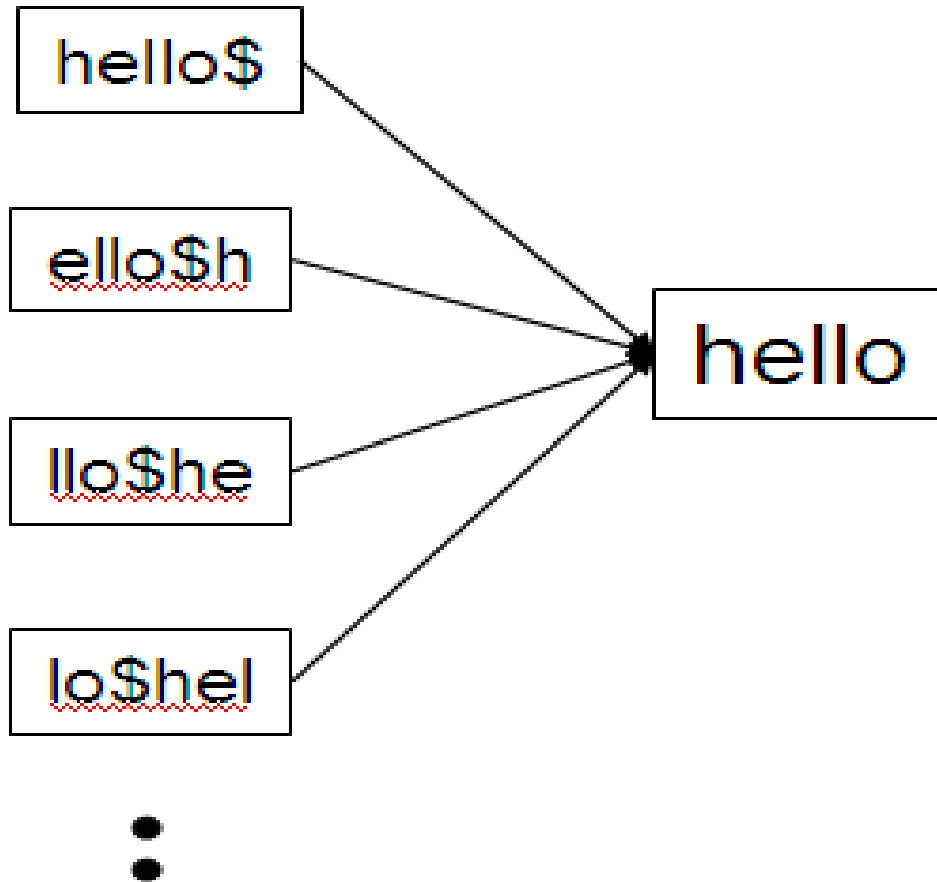
lo\$hel

o\$hell

\$hello

to the B-tree where \$ is a special symbol

Term mapping



Contd..

- *hello*
- Queries
- For X , look up $X\$$ For X^* , look up $\$X^*$ For $*X$, look up $X\* For $*X^*$, look up X^*
- For X^*Y , look up $Y\$X^*$
- Example: For $hel*o$, look up $o\$hel^*$
- Permuterm index would better be called a [permuterm tree](#).

Thank You