

Intro. to Search Engine

28/11/2017

Normal Search vs. Search Engine

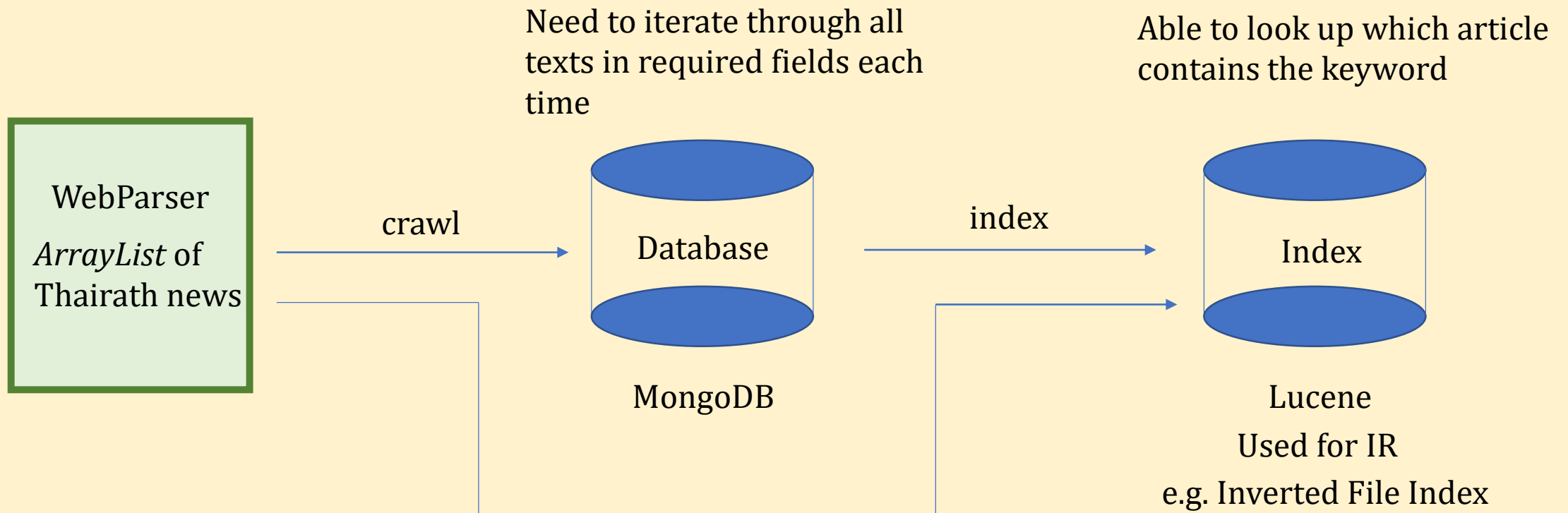
Finer
differences



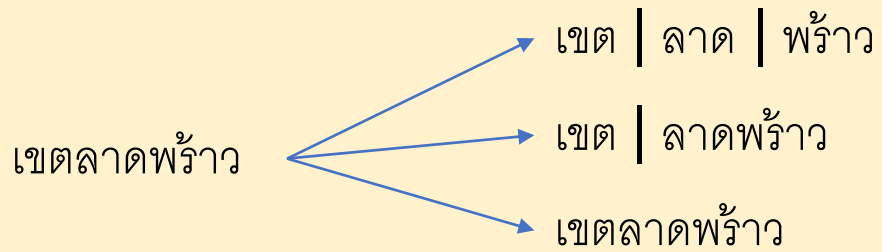
Table	Search Engine
Database	Index
Table/Column attribute	Index
Row	Document
Attribute/Column	Field
Any data types	String only

Stored in "Big Table"

Full-text Search: Database vs. Index



Tokenizing methods are specific for different purposes.



Which to choose for tokenizing?

ไปห้ามเหสี

Longest
matching

ไป | ห้าม | เห | สี

Maximal
matching

ไป | หา | มเหสี

Possibly....

ไป | หา | ม | เหสี

What if? %ตลาด%

เขตลาดพร้าว will be recognized as a hit for string matching.

Which one is the best?
...depends on applications!...