

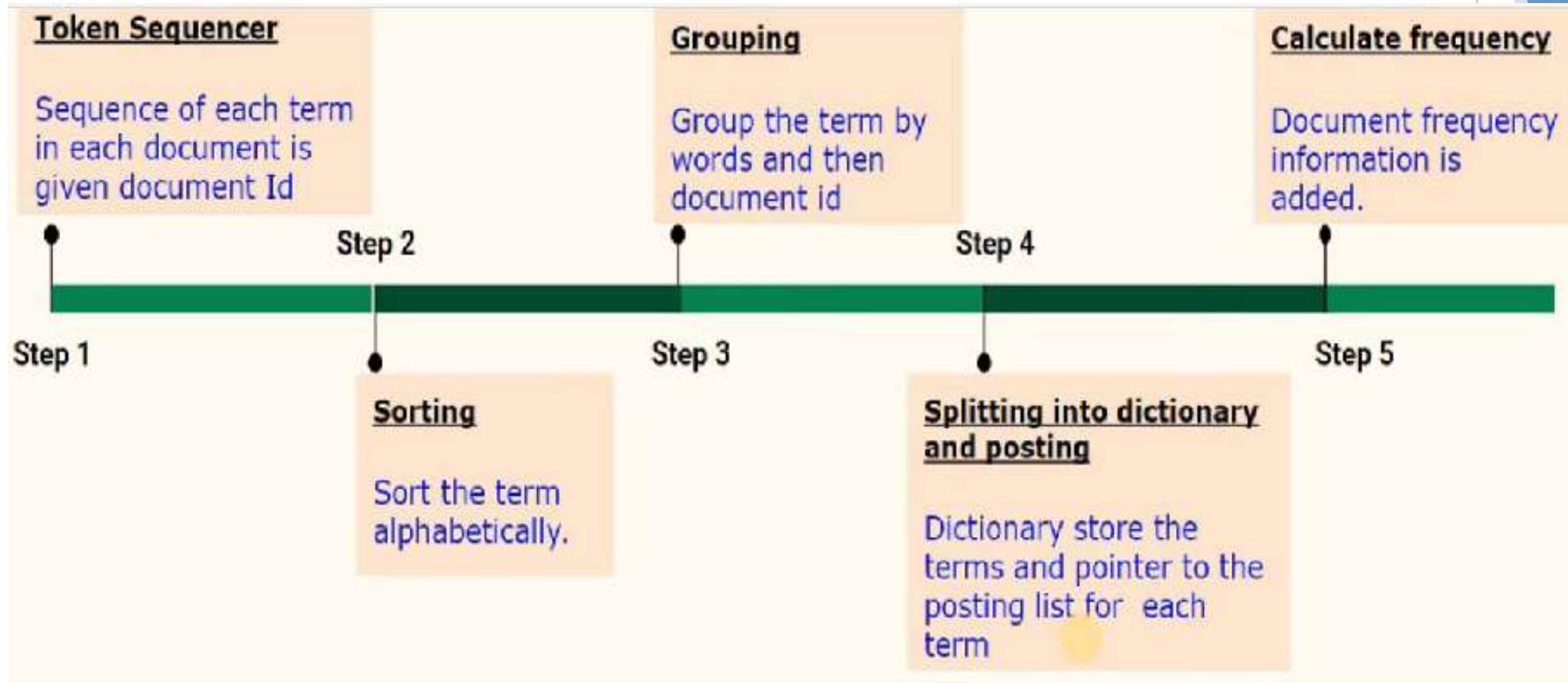
Inverted Index and Boolean Query

UNIT-1

What is an Inverted index?

- ▶ An inverted index is an index data structure storing a mapping from content such as words or numbers, to its location in a document or a set of documents.
- ▶ It is a hashmap like data structure that directs you from a word to a document or a web page.

Building an Inverted Index



Doc1 :

Ram was average student in study but once he became friend of Shyam he started attending lecture regularly.

Doc2:

Shyam was enthusiastic student. He had helped his friend Ram a lot. Now Ram and Shyam are good friends.

Building an inverted index - Example

Term	DocID
Ram	1
was	1
average	1
student	1
in	1
study	1
but	1
once	1
he	1
became	1
friend	1
of	1
Shyam	1
he	1
started	1
attending	1
lecture	1
regularly	1

Term	DocID
Shyam	2
was	2
very	2
enthusiastic	2
student	2
He	2
had	2
helped	2
his	2
friend	2
Ram	2
a	2
lot	2
Now	2
Ram	2
and	2
Shyam	2
are	2
good	2
friend	2



Term	DocID
a	2
and	2
are	2
attending	1
average	1
became	1
but	1
enthusiastic	2
friend	1
friend	2
friend	2
good	2
had	2
he	1
he	1
He	2
helped	2
his	2
in	1
in	1
lecture	1

Term	DocID
lecture	1
lot	2
lot	2
Now	2
Now	2
of	1
once	1
Ram	1
Ram	2
Ram	2
regularly	1
Shyam	1
Shyam	2
Shyam	2
started	1
student	1
student	2
study	1
very	2
was	1
was	2

Building an inverted index - Example

Term	DocID
a	2
and	2
are	2
attending	1
average	1
became	1
but	1
enthusiastic	2
friend	1
friend	2
friend	2
good	2
had	2
he	1
he	1
He	2
helped	2
his	2
in	1
lecture	1
lot	2

Term	DocID
Now	2
in	1
lecture	1
lot	2
Now	2
of	1
once	1
Ram	1
Ram	2
Ram	2
regularly	1
Shyam	1
Shyam	2
Shyam	2
started	1
student	1
student	2
study	1
very	2
was	1
was	2

Term	Doc. Frequency
a	1
and	1
are	1
attending	1
average	1
became	1
but	1
enthusiastic	1
friend	2
good	1
had	1
he	1
He	1
helped	1
his	1
in	1
lecture	1
lot	1
Now	1
of	1
once	1
Ram	2
regularly	1
Shyam	2
started	1
student	2
very	1
was	2

Building an inverted index - Example

Term	DocID	Term	DocID
a	2	Now	2
and	2	in	1
are	2	lecture	1
attending	1	lot	2
average	1	Now	2
became	1	of	1
but	1	once	1
enthusiastic	2	Ram	1
friend	1	Ram	2
friend	2	Ram	2
friend	2	regularly	1
good	2	Shyam	1
had	2	Shyam	2
he	1	Shyam	2
he	1	started	1
He	2	student	1
helped	2	student	2
his	2	study	1
in	1	very	2
lecture	1	was	1
lot	2	was	2

Term	Doc. Frequency	Posting List
a	1	2
and	1	2
are	1	2
attending	1	1
average	1	1
became	1	1
but	1	1
enthusiastic	1	2
friend	2	1 → 2
good	1	2
had	1	1
he	1	1
He	1	2
helped	1	2
his	1	2
in	1	1
lecture	1	1
lot	1	2
Now	1	2
of	1	1
once	1	1
Ram	2	1 → 2
regularly	1	1 → 2
Shyam	2	1 → 2
started	1	1
student	2	1 → 2
very	1	2
was	2	1 → 2