Indexes in BaseX

Dimitar Popov

University of Konstanz

Types of Indexes: Overview

Structure Indexes

- Document index

- Tag and attribute name index

- Path summary index

Data Indexes

- Text and attribute value index

- Full-text index

Statistics

Document Index

Purpose

- Document path storage

- Fast access to documents by name

Example

- List documents in a folder:

  `db:list('db', 'mydir')`

- Retrieve document(s):

  `collection('db/mydir')`

Structure

- Document paths in PRE-order

- Index of sorted document paths

- In-memory

Tag- and Attribute-Name Index

Purpose

- Name storage

- Fast name look-up

- Statistics

Example

- Name tests:

  `//country/citys`

- Statistics:

  `count(//city)`

Structure

- Hash index

- In-memory

Path-Tree Index

Purpose

- Typing suggestions

- Path optimizations

- Statistics

Example

- Re-writing queries

  ```
  //country/city
  /mondial/country/city
  ```

Structure

- Tree representing the XML structure

- In-memory

Text and Attribute Values Index

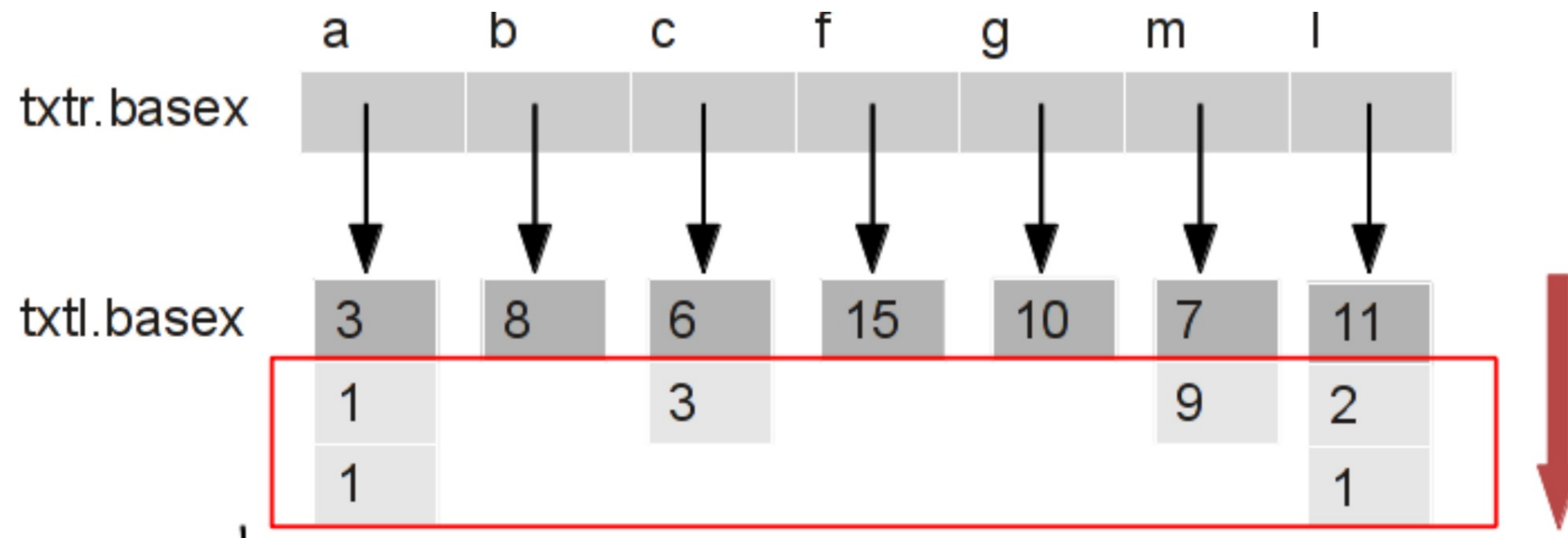## Purpose

- Search text nodes

- Attribute values

Example

- Re-write queries

```
//city[./name/text() = 'Prague']
db:text("fb", "Prague")/parent::*:name /parent::*:city
```

Text and Attribute Values Index

Structure

Full-Text Index

Purpose

- Full-text queries

- Fuzzy search using Levenshtein

Example

- Re-writing queries

```
//city/description[text() contains text 'Prague']
db:fulltext("fb", "Prague")
/parent::*:description[parent::*:city]
```

Index Updates

Prerequisites

- Using stable node identifiers

- Mapping between PRE and ID

Disadvantages

- Update operations are slower

- Performance may degrade (no more PREs in index)

Advantages

- No manual recreation

Index Updates: Performance

Database:	xmark111mb
Size:	107 MiB
Input Size:	111 MiB
Nodes:	3221926
Text Index Entries:	297743
Text Index Size:	5 MiB
Attr. Index Entries:	71019
Attr. Index Size:	2 MiB

	Off	Total	Attr.	Text	#Text	#Attr
Big Insert	3232,13	5823,57	653,59	2866,55	213794	48840
Big Delete	14,23	3102,00	321,62	2378,61	213794	48840
Small Insert	165,07	192,83	3,96	27,07	11	11
Small Delete	22,01	245,42	136,17	213,89	11	11

Questions?

Thank you!