

CS347

Lecture 2

April 9, 2001

©Prabhakar Raghavan

Today's topics

- Inverted index storage
 - Compressing dictionaries into memory
- Processing Boolean queries
 - Optimizing term processing
 - Skip list encoding
- Wild-card queries
- Positional/phrase queries
- Evaluating IR systems

Recall dictionary and postings files

| Term | Doc # | Freq |
|-----------|-------|------|
| ambitious | 2 | 1 |
| be | 2 | 1 |
| brutus | 1 | 1 |
| brutus | 2 | 1 |
| capitol | 1 | 1 |
| caesar | 1 | 1 |
| caesar | 2 | 2 |
| did | 1 | 1 |
| enact | 1 | 1 |
| hath | 2 | 1 |
| I | 1 | 2 |
| i' | 1 | 1 |
| it | 2 | 1 |
| julius | 1 | 1 |
| killed | 1 | 2 |
| let | 2 | 1 |
| me | 1 | 1 |
| noble | 2 | 1 |
| so | 2 | 1 |
| the | 1 | 1 |
| the | 2 | 1 |
| told | 2 | 1 |
| you | 2 | 1 |
| was | 1 | 1 |
| was | 2 | 1 |
| with | 2 | 1 |



| Term | N docs | Tot Freq |
|-----------|--------|----------|
| ambitious | 1 | 1 |
| be | 1 | 1 |
| brutus | 2 | 2 |
| capitol | 1 | 1 |
| caesar | 2 | 3 |
| did | 1 | 1 |
| enact | 1 | 1 |
| hath | 1 | 1 |
| I | 1 | 2 |
| i' | 1 | 1 |
| it | 1 | 1 |
| julius | 1 | 1 |
| killed | 1 | 2 |
| let | 1 | 1 |
| me | 1 | 1 |
| noble | 1 | 1 |
| so | 1 | 1 |
| the | 2 | 2 |
| told | 1 | 1 |
| you | 1 | 1 |
| was | 2 | 2 |
| with | 1 | 1 |

↑
In memory

Gap-encoded,
on disk

| Doc # | Freq |
|-------|------|
| 2 | 1 |
| 2 | 1 |
| 1 | 1 |
| 2 | 1 |
| 1 | 1 |
| 1 | 1 |
| 2 | 2 |
| 1 | 1 |
| 1 | 1 |
| 2 | 1 |
| 1 | 2 |
| 1 | 1 |
| 2 | 1 |
| 1 | 1 |
| 1 | 1 |
| 2 | 1 |
| 1 | 1 |
| 2 | 1 |
| 1 | 1 |
| 2 | 1 |
| 1 | 1 |
| 2 | 1 |
| 2 | 1 |
| 1 | 1 |
| 2 | 1 |
| 2 | 1 |

