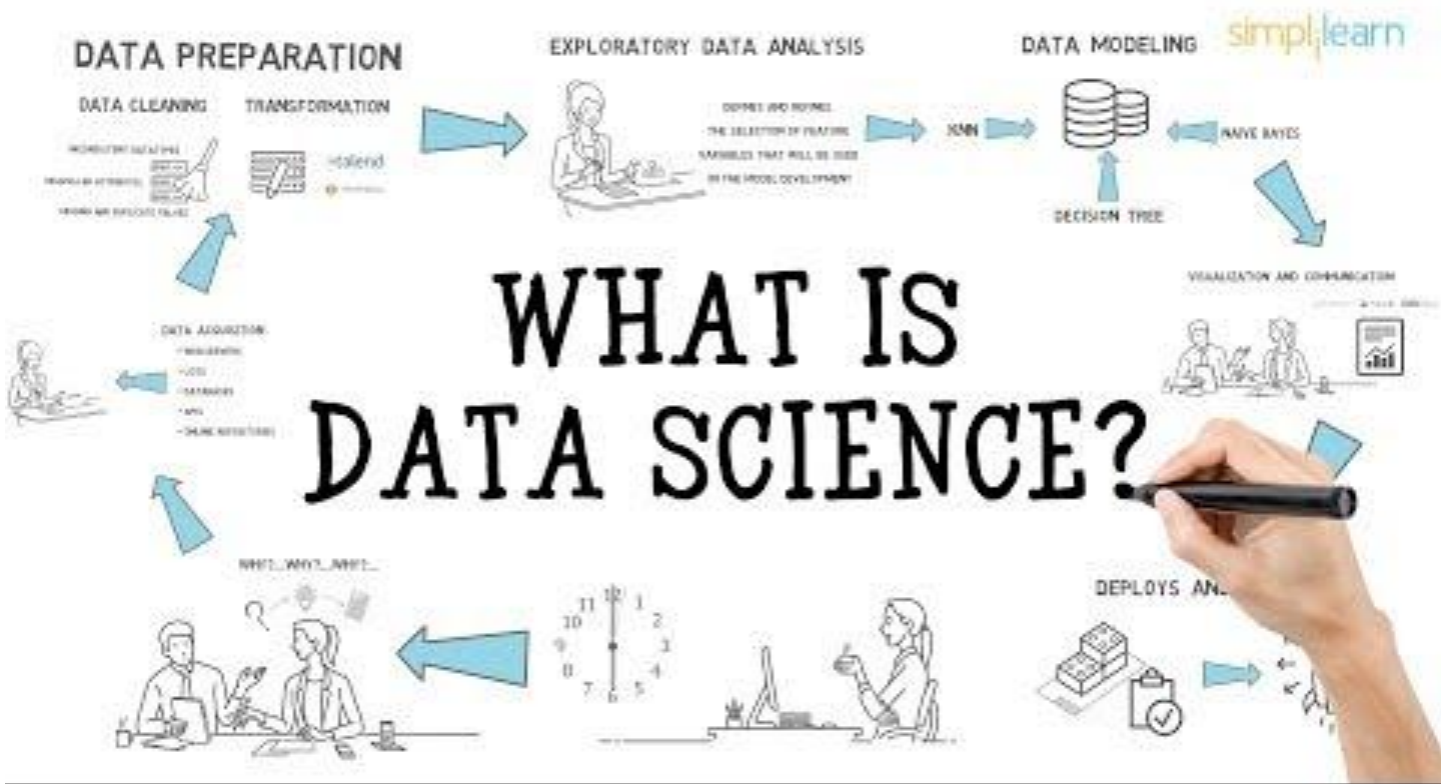
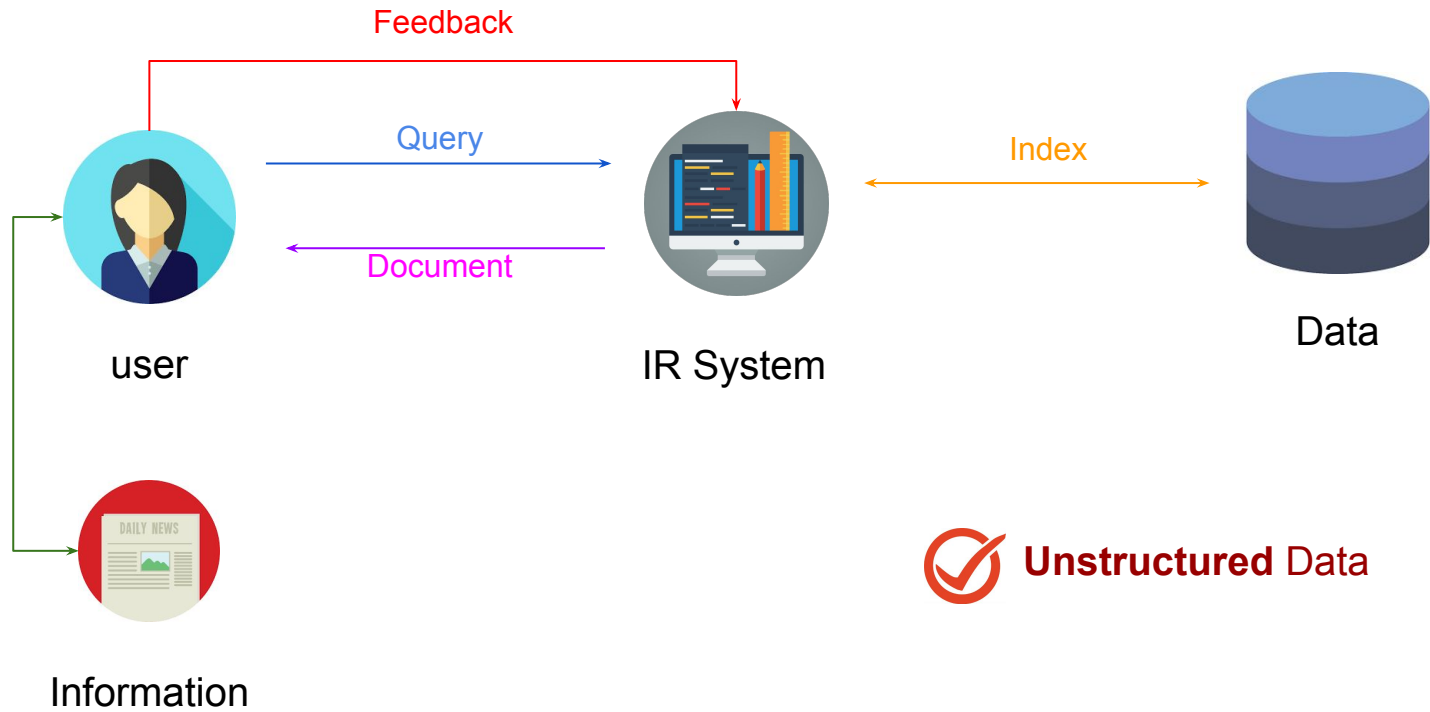


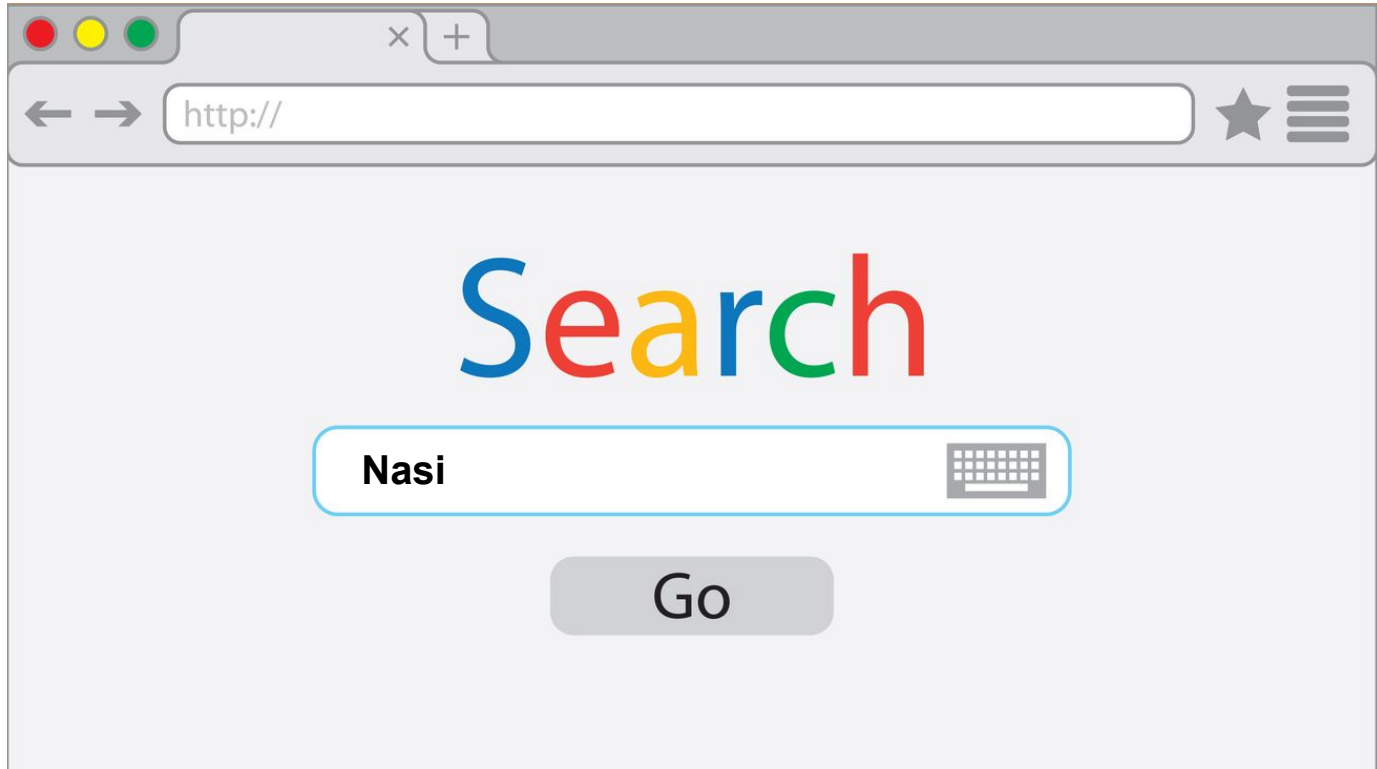
Pertemuan 2



Information retrieval



Intro Indexing



nasi



Mau Masak Nasi Pulen dan Harum? Perhatikan 10 Hal ...

<https://food.detik.com> › info-kuliner › mau-masak-nasi-... ▾ Translate this page

Jul 11, 2019 - Masak **nasi** terlihat gampang. Padahal untuk masak **nasi** yang pas pulennya dan harum, ada tips penting.

Bosan Nasi Putih? Coba Masak Nasi Pakai Milk Tea dan Boba

<https://food.detik.com> › info-kuliner › bosan-nasi-putih-... ▾ Translate this page

Jun 26, 2019 - Dilansir dari SoraNews 24 (26/6) milk tea ternyata bisa dimasak dengan **nasi**. Percobaan unik sekaligus nyentrik ini ternyata menghasilkan ...

10 Jenis Olahan Nasi di Jakarta, Kamu Sudah Pernah Coba ...

<https://travel.kompas.com> › Travel › Food Story ▾ Translate this page

Jun 22, 2019 - Tak hanya **nasi** putih, Jakarta memiliki beragam olahan **nasi** yang patut kamu coba. Halaman all.

Jangan Salah, Rutin Makan Nasi Putih Tingkatkan Risiko ...

<https://sains.kompas.com> › Sains › Kita ▾ Translate this page

Jun 21, 2019 - Tidak perlu berhenti makan **nasi** putih. Mulailah dengan cara ini. Halaman all.

Resep Nasi Masako® - Dapur Umami

<https://www.dapurumami.com> › resep › nasi-masako-Ve... ▾ Translate this page

Ada yang bilang, hari belum lengkap kalau belum makan **nasi**, apakah ini berlaku juga untuk Mama sekeluarga? Cobain yuk, ide baru makan **nasi**, yaitu **Nasi** ...

Berapa
Dokumen ?

Inverted Index

- ✓ Merupakan **Struktur data utama** pada sistem IR
- ✓ Disebut juga **Inverted File**
- ✓ Berisi **List** yang memuat **term**
- ✓ pada setiap List terdapat **deretan dokumen** yang memuat term tsb

Index Dokumen dilakukan untuk menghindari pencarian dokumen secara linier

Inverted Index

inverted Index @ index yg memetakan keyword / term ke daftar dokumen

Dictionary @ kumpulan term / keyword

Posting List @ Daftar dokumen yang disosialisasikan dg keyword yg di berikan

Inverted Index

Posting List

Document 1

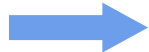
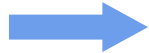
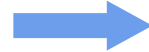
The bright blue butterfly hangs on the breeze.

Document 2

It's best to forget the great sky and to retire from every wind.

Document 3

Under blue sky, in bright sunlight, one need not search around.



inverted Index

| ID | Term | Document : position |
|----|-----------|---------------------|
| 1 | best | 2 : 3 |
| 2 | blue | 1 : 3, 3 : 2 |
| 3 | bright | 1 : 2, 3 : 5 |
| 4 | butterfly | 1 : 4 |
| 5 | breeze | 1 : 8 |
| 6 | forget | 2 : 5 |
| 7 | great | 2 : 7 |
| 8 | hangs | 1 : 5 |
| 9 | needs | 3 : 8 |
| 10 | retire | 2 : 11 |
| 11 | search | 3 : 10 |
| 12 | sky | 2 : 8, 3 : 3 |
| 13 | wind | 2 : 14 |

Langkah2 Inverted Index

Fetching the Document

Tokenization

Stop Word Remover

Normalization

Record Document IDs

Tokenization

memecah kalimat menjadi bagian-bagian yang disebut **token** dan **menghilangkan** bagian tertentu seperti **tanda baca**

Input : **Saya makan bakso di pasar.**

Output :

Saya

makan

bakso

di

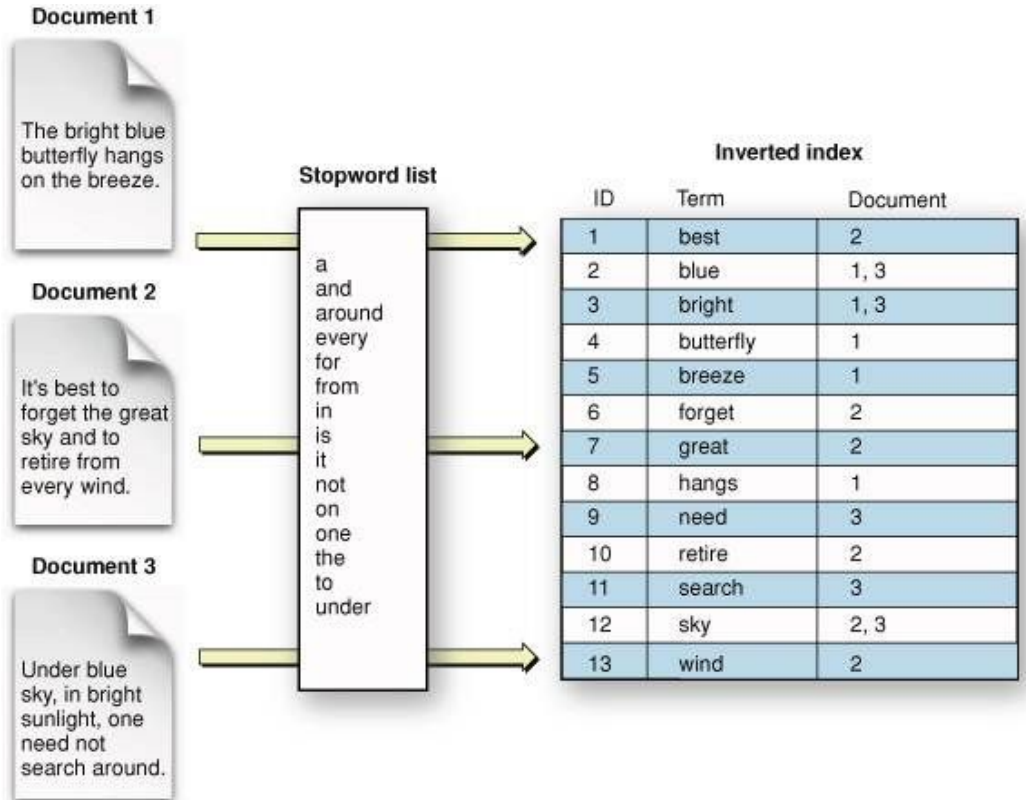
pasar

Stopword

kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna

Stopword bahasa Indonesia

“yang”, “dan”, “di”, “dari”



Normalization

merupakan upaya “menyamakan bentuk”

Beberapa metodenya

menghilangkan tanda baca untuk singkatan

U.S.A → USA

Case Folding

INDONESIA → indonesia

Stemming

Lemmatization

Stemming

merupakan upaya mengubah sebuah kata ke bentuk asal “**Kata Dasar**”

menghilangkan imbuan dari sebuah kata (**awalan, akhiran, sisipan**)

membawa

dibawa

terbawa



bawa

Lemmatization

merupakan upaya mengubah sebuah kata ke bentuk asal “**Kata Dasar**”

dengan menggunakan bantuan **Kamus / Dictionary** sebagai acuan

meny..... berasal dari **men + s.....**

menyiram → **SIRAM**

menyapu → **SAPU**

menyalin → **SALIN**

Langkah2 pencarian

blue sky



1

Tokenizer

✓ blue

✓ sky

Inverted index

| ID | Term | Document |
|----|-----------|----------|
| 1 | best | 2 |
| 2 | blue | 1, 3 |
| 3 | bright | 1, 3 |
| 4 | butterfly | 1 |
| 5 | breeze | 1 |
| 6 | forget | 2 |
| 7 | great | 2 |
| 8 | hangs | 1 |
| 9 | need | 3 |
| 10 | retire | 2 |
| 11 | search | 3 |
| 12 | sky | 2, 3 |
| 13 | wind | 2 |

Langkah2 pencarian

2

cari tiap token dalam inverted index

blue

1

3

sky

2

3

3

Dilakukan Marge

blue sky

3

Document 3

Under blue
sky, in bright
sunlight, one
need not
search around.

Contoh Kasus

Persebaya Surabaya harus puas dengan hasil seri 1-1 saat menjamu Persik Kediri dalam laga pembuka Shopee Liga 1 2020.

Skor imbang melawan Persik Kediri tidak diharapkan, tapi bukan berarti performa Persebaya buruk

Persebaya membuka musim 2020 dengan hasil seri. Menurut Aji Santoso timnya kurang beruntung

Persebaya harus berbagi angka dengan Persik Kediri pada laga pembuka Shopee Liga 1 2020