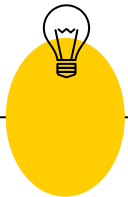


# Sistem Temu Kembali Informasi

“Web Mining”





## Buku Penunjang & Literatur





# Course **Outline**





# Pengenalan



Kumpulan file yang terkait dalam web server.

Merupakan repository terbesar data terdiri atas HTML, XML, format teks, etc.

Penerapan teknik data mining untuk mengekstrak pengetahuan dari data web.



# Tantangan Pengolahan Data Web

## CHALLENGES

Terlalu  
besar untuk  
efektif data  
mining.

Terlalu  
rumit dan  
dinamis.

Tidak  
spesifik  
untuk  
domain.

Memiliki  
segalanya.

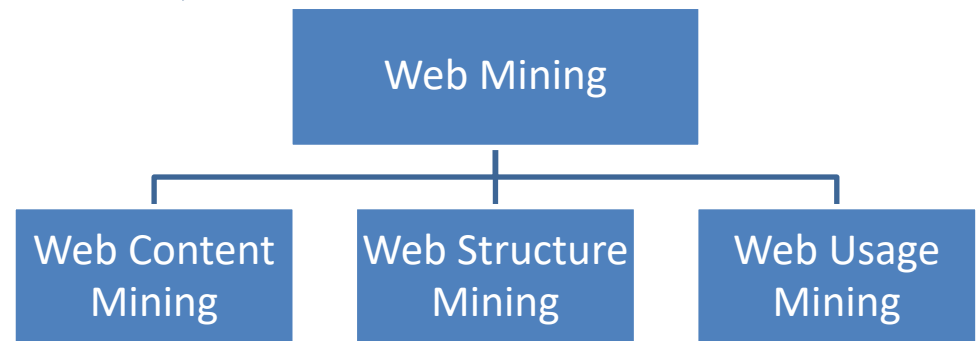


# Konsep Web Mining

## WEB MINING



proses menemukan hubungan intrinsik dari data web (tekstual, linkage, atau penggunaan).





# WEB CONTENT MINING





# WEB CONTENT MINING

CONTENT  
MINE



proses penggalian informasi yang berguna dari isi dokumen Web.



Data konten sesuai dengan kumpulan fakta dari halaman Web yang dirancang untuk disampaikan kepada pengguna.



Terdiri atas teks, gambar, audio, video, atau catatan terstruktur seperti daftar (*lists*) dan tabel.





# Penerapan Web Content Mining





# WEB STRUCTURE MINING





# Web Structure Mining

Web Structure Mining terdiri dari halaman Web sebagai node, dan hyperlink sebagai tepi menghubungkan antara dua halaman yang terkait.



Web Structure dapat melakukan proses penemuan struktur informasi dari Web.



Dapat dilakukan baik pada (intra-halaman) tingkat dokumen atau di (inter-halaman) tingkat hyperlink. Penelitian di tingkat hyperlink juga disebut ***Hyperlink Analysis***.



- Representasi yang mendasari dalam membentuk dasar untuk melaksanakan aplikasi tugas tertentu.



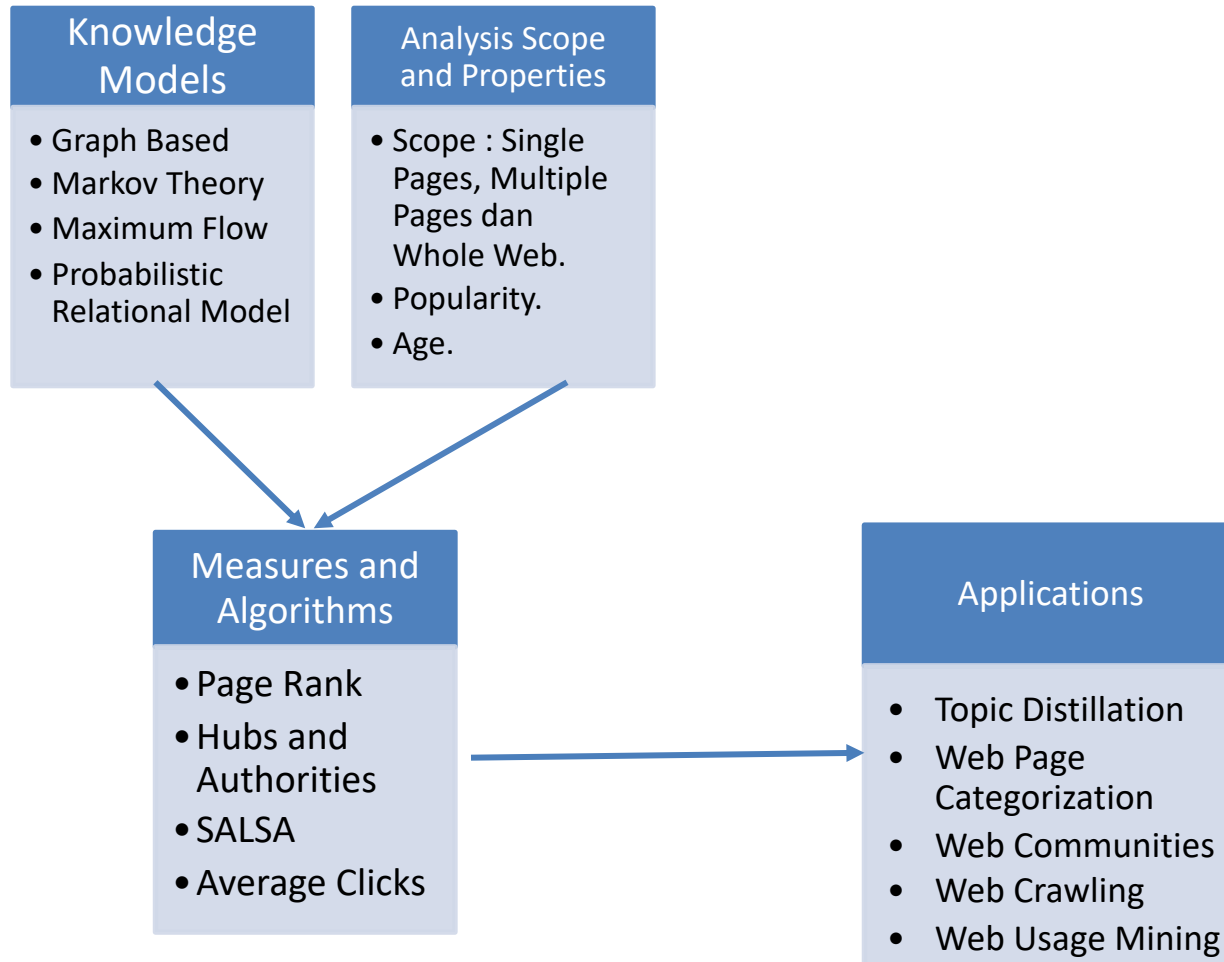
- Ruang lingkup analisis yang menentukan jika tugas relevan dengan node tunggal atau set node atau seluruh grafik.
- Properti adalah karakteristik dari node tunggal atau set node atau seluruh web.



- Pengukuran standar untuk properti seperti kualitas, relevansi atau jarak antara node.
- Algoritma dirancang untuk perhitungan efisiensi pengukuran tersebut.

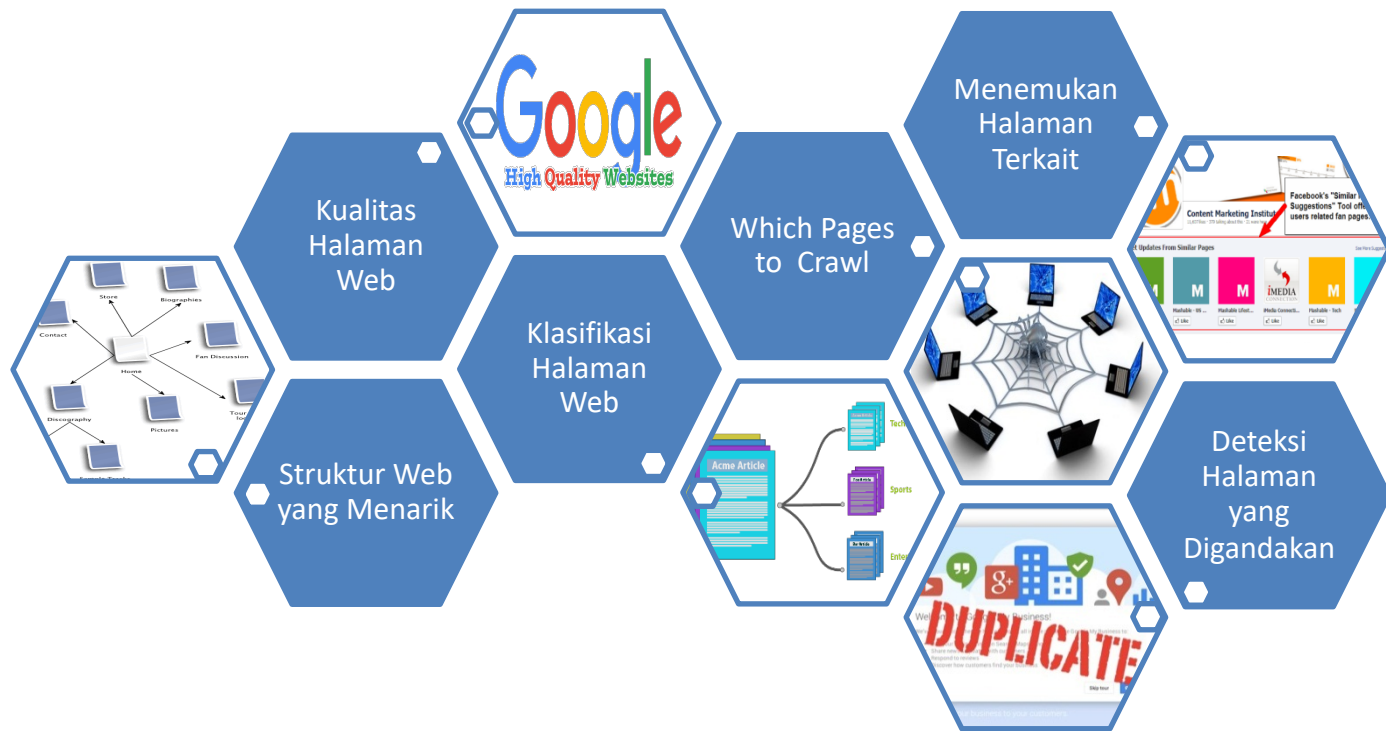


# Teknik Hyperlink Analysis (2)





# Penerapan Web Structure Mining





# WEB USAGE MINING





# Web Usage Mining

## Web Usage Mining

Ekstraksi informasi dari data yang dihasilkan melalui kunjungan halaman Web dan transaksi.

Data yang disimpan dalam log akses server, *referrer logs*, log agen, dan *client-side cookies*.

Karakteristik pengguna dan profil penggunaan, metadata seperti atribut halaman, atribut konten, dan data penggunaan.

Data clickstream dan analisis clickstream.







# Penerapan Web Usage Mining

Menentukan nilai  
*lifetime* klien.

Desain strategi cross-  
marketing seluruh produk.

Evaluasi Promosi  
Kampanye

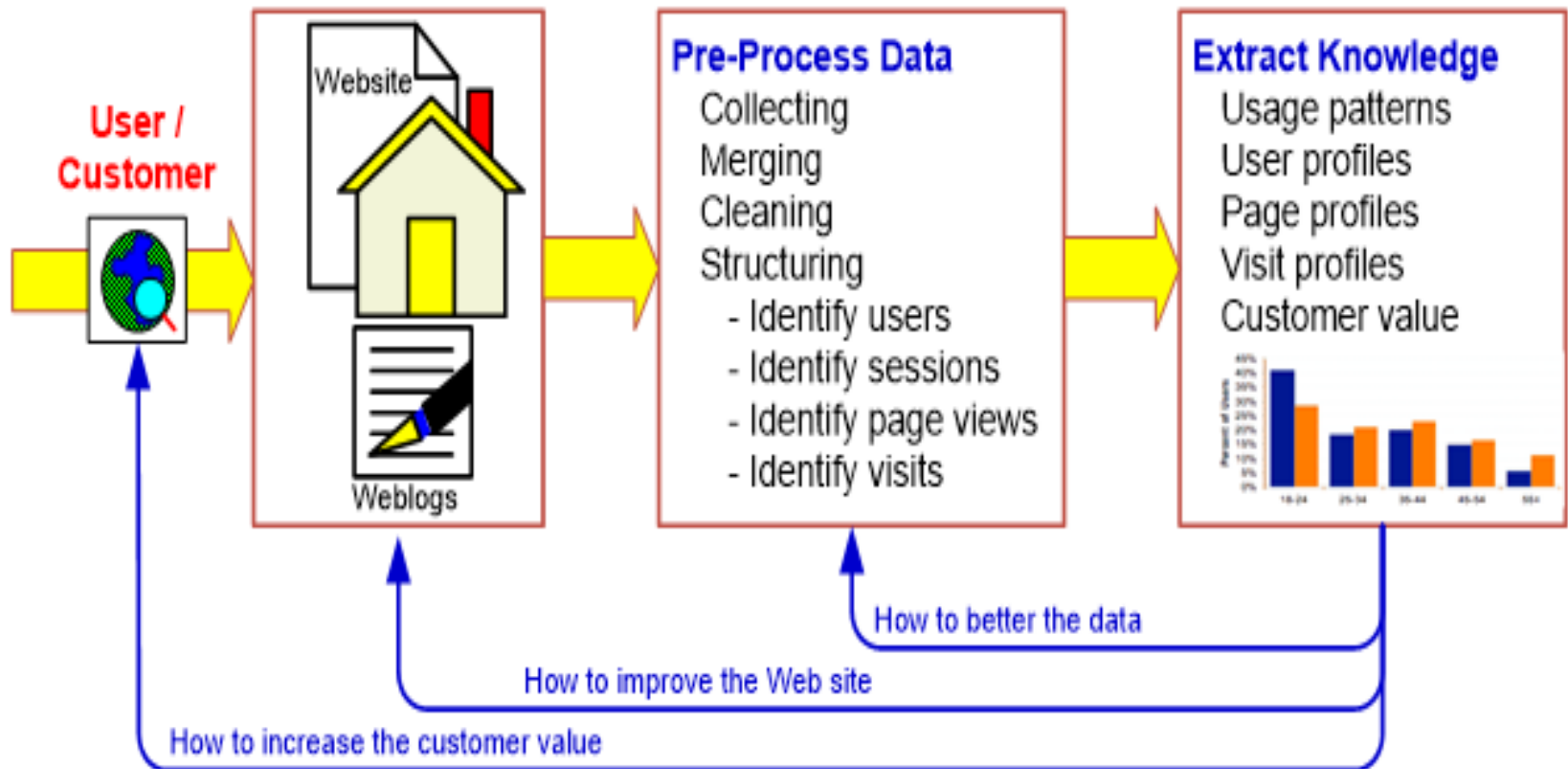
Menargetkan iklan elektronik dan kupon di kelompok  
pengguna berdasarkan pola akses pengguna.

Memprediksi perilaku pengguna berdasarkan  
aturan dipelajari sebelumnya dan profil pengguna.

Menampilkan informasi dinamis untuk  
pengguna berdasarkan minat dan profil mereka.



# Web Usage Mining (Analisis Clickstream)

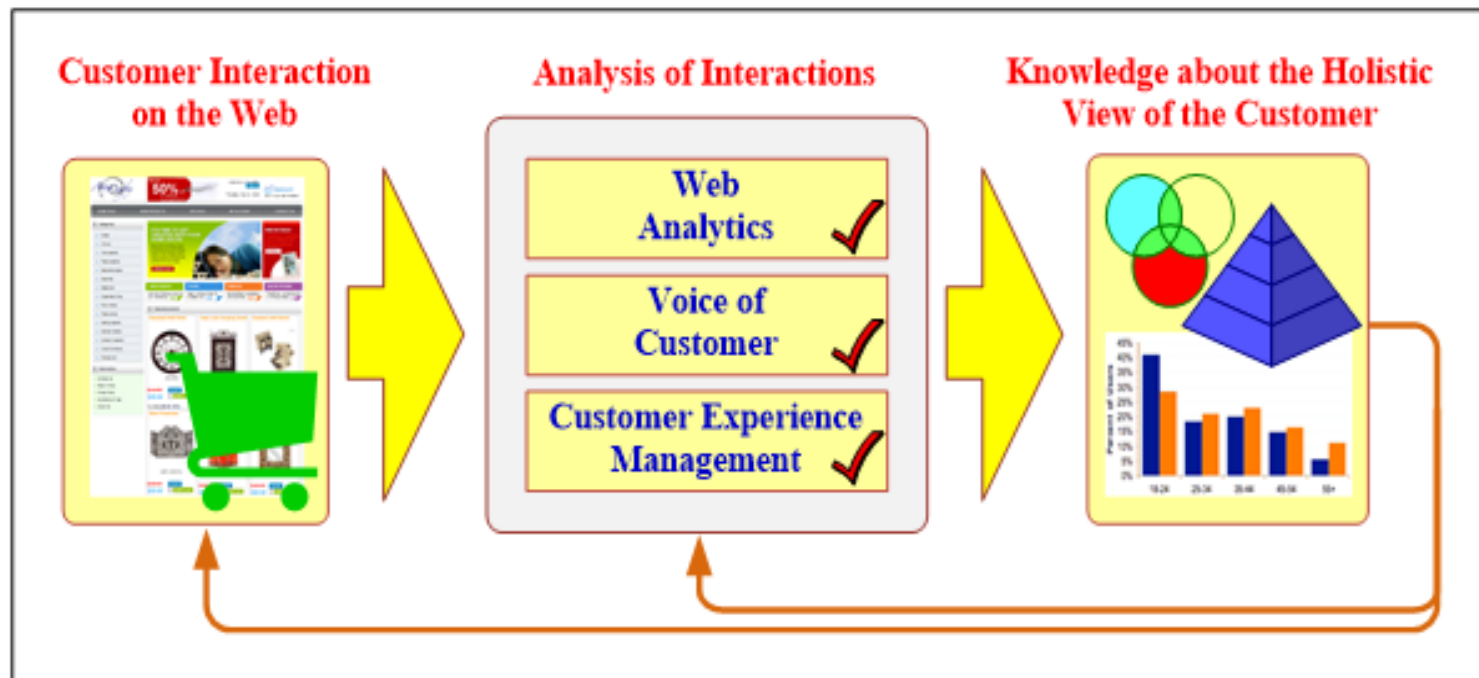




# Web Mining Success Stories

Amazon.com, Ask.com, Scholastic .com, etc . . .

Website Optimization Ecosystem



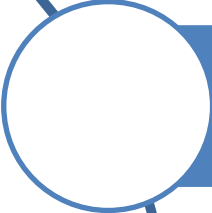


# Web Mining Tools

No	Nama Produk	URL
1	Rapidminer	rapidminer.com
2	Angoss Knowledge Web Miner	angosss.com
3	ClickTracks	clicktracks.com
4	LiveStats from DeepMetrix	deepmetrix.com
5	Megaputer Web Analyst	megaputer.com
6	MicroStrategy Web Traffic Analysis	microstrategy.com
7	SAS Web Analytics	sas.com
8	SPSS Web Mining for Clementine	spss.com
9	WebTrends	webtrends.com
10	XML Miner	scientio.com



# Referensi



Efraim Turban, Ramesh Sharda and Dursun Delen. *Decision Support and Business Intelligence Systems* 9th Ed. Prentice Hall. 2011.



Min-Yuh Day. Introduction to Web Mining – Lecture Notes. 2012.



N Girija. An Overview – Web Mining. The ICFAI University Press.



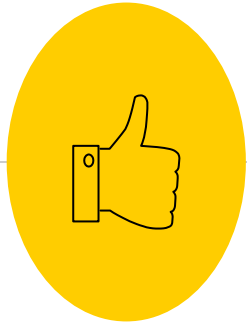
## Kesimpulan & Review

- Web Mining : proses menemukan hubungan intrinsik dari data web (tekstual, linkage, atau penggunaan).
- Web Mining dibedakan menjadi tiga bagian yaitu :
  1. Web Content Mining
  2. Web Structure Mining
  3. Web Usage Mining



## Kuis (Latihan Soal)

- Carilah paper tentang Sistem Temu Kembali Informasi (STKI) minimal 3 paper yang berkaitan dengan materi **Web Mining**.
- Rangkumlah menjadi artikel baru tentang review masalah diatas (**minimal 2 halaman kertas A4**).



# Thanks!

*Any questions ?*