

# Pengantar Web Mining

## Pencarian dan Penambangan Web

Prodi Teknik Informatika

Universitas Trunojoyo Madura

2024

# Pengantar

- Kuliah ini memperkenalkan konsep dasar dan teknologi dari web mining.
- Topik terdiri
  - Pengantar Web Mining,
  - Pembelajaran Terawasi (Supervised Learning)
  - Pembelajaran Tak terawasi (Unsupervised Learning)
  - Analisa Jejaring Sosial (Social Network Analysis)
  - Web Crawling,
  - Ekstraksi data terstruktur (Structured Data Extraction)
  - Analisa Sentimen dan Menambang Opini (Opinion Mining and Sentiment Analysis)
  - Web Usage Mining.

# Tujuan

- Mahasiswa akan dapat memahami dan menggunakan konsep dasar dan teknologi web mining.
- Mahasiswa akan dapat melakukan penelitian sistem informasi dalam konteks web mining.

# Web Mining

- Penambangan web (Web mining)
  - Menemukan informasi yang berguna atau pengetahuan dari
    - Struktur hyperlink laman ( **Web hyperlink structure**),
    - Isi laman (**page content**)
    - **usage data**.
- Tiga macam tugas web mining
  - Web structure mining
  - Web content mining
  - Web usage mining

# Penambangan teks (Text Mining)

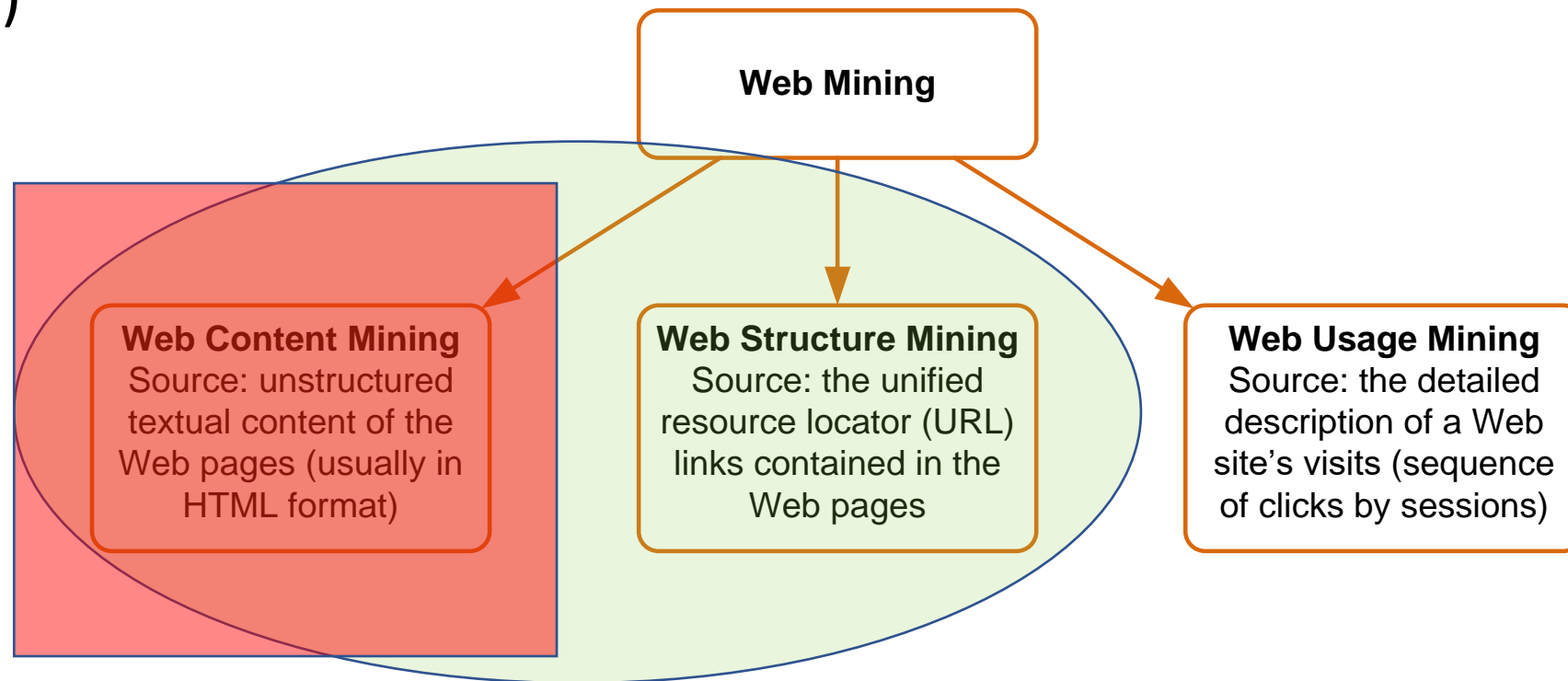
- Penambangan teks (Text mining) (text data mining)
  - Proses mendapatkan informasi yang berkualitas dari teks
- Macam macam tugas text mining
  - text categorization/classification
  - text clustering
  - concept/entity extraction/ topic modelling
  - sentiment analysis
  - document summarization

# Penjelasan Web Mining

- Web adalah pangkalan data yang sangat besar
- Datanya dalam format HTML, XML, text
- Tantangan pemrosesan data Web)
  - Web adalah sangat besar untuk dilakukan penambahan data
  - Web sangat komplek
  - Web sangat dinamis
  - Web bukan domain yang spesifik
  - Web adalah segalanya

# Web Mining

- Web mining (atau Web data mining) adalah proses untuk menemukan keterkaitan yang tersembunyi dari data web (**textual**, linkage, atau usage)



# Web Content

- Menambang isi tektual pada Web
- Pengumpulan data dengan Web crawler



# Web Usage Mining

- Extraction of information from data generated through Web page visits and transactions...
  - data stored in server access logs, referrer logs, agent logs, and client-side cookies
  - user characteristics and usage profiles
  - metadata, such as page attributes, content attributes, and usage data
- Clickstream data
- Clickstream analysis

# Aplikasi Text Mining

- Information extraction
- Topic modelling
- Summarization
- Categorization
- Clustering

# Terminology

- Unstructured or semistructured data
- Corpus (and corpora)
- Terms
- Concepts
- Stemming
- Stop words (and include words)
- Tokenizing

# Web Usage Mining

- Aplikasi Web usage mining
  - Memprediksi karakteristik user
  - Sistem Rekomendasi / dynamic information berdasarkan profile dan historis dari user
  - Personalisasi Konten Web:
    - Menyesuaikan konten web berdasarkan preferensi dan perilaku pengguna untuk memberikan pengalaman yang lebih relevan dan menarik1.
  - Analisis Perilaku Pengguna:
    - Memahami bagaimana pengguna berinteraksi dengan situs web untuk mengidentifikasi pola dan tren yang dapat digunakan untuk meningkatkan desain situs web dan strategi pemasaran2.
  - Optimasi Mesin Pencari (SEO):
    - Menganalisis kueri mesin pencari dan halaman hasil mesin pencari (SERP) untuk meningkatkan visibilitas situs web dalam hasil pencarian dan meningkatkan lalu lintas ke situs web2.

# Web Usage Mining

- Aplikasi Web usage mining
  - Deteksi Penipuan:
    - Mengidentifikasi aktivitas penipuan di situs web untuk mencegah penipuan finansial, pencurian identitas, dan jenis penipuan online lainnya2.
  - Analisis Sentimen:
    - Menganalisis data media sosial untuk mengekstrak sentimen dari posting, komentar, dan ulasan guna memahami perasaan pelanggan terhadap produk dan layanan2.
  - Analisis Konten Web:
    - Menganalisis konten web untuk mengekstrak informasi berharga seperti kata kunci, topik, dan tema yang dapat digunakan untuk meningkatkan relevansi konten web dan mengoptimalkan peringkat mesin pencari

# Web Usage Mining (clickstream analysis)

