

Data Mining = การทำเหมืองข้อมูล เพื่อค้นหาความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลชุดนั้น / จัดหมวดหมู่ในเชิงการกระทำที่น่าจะได้จาก ข้อเคมคือ knowledge discovery (mining) in Database (KDD) → การสกัดความรู้จากฐานข้อมูลขนาดใหญ่แล้ว ส่องหา knowledge extraction, data Analysis, data archeology, business intelligence (การสกัดความรู้จาก Data เพื่อได้ข้อสันนิษฐานธุรกิจ)



♥ Web Mining Framework.

- Data Cleaning
- Data integration from multiple sources
- Warehousing the data เก็บไว้เพื่อให้นักวิจัยไม่ต้องไปหาก็ได้
- Data Cube construction
- Data selection for Data mining
- Data mining
- Presentation of mining results แสดงองค์ความรู้ให้นักวิจัยเห็นได้
- Patterns and knowledge to be used or stored into knowledge-base

♥ KDD Process: A View from ML and Statistics

Input Data → Data Pre-processing → Data Mining → Post-processing → Pattern Information knowledge

Data Integration
Normalization
Feature selection
Dimension reduction

Pattern discovery → หาแบบที่ซ่อนอยู่ในข้อมูล
Classification → จัดแนกข้อมูล
Clustering → แบ่งกลุ่มข้อมูล
Outlier analysis

Pattern evaluation
Pattern selection
Pattern interpretation
Pattern visualization

★ Row: Record, Data point: ข้อมูลแต่ละตัว

★ เทคนิค Data Mining

1. Pattern Discovery มาหาคำตอบแบบที่เรากำลังมองหาในข้อมูล ชื่อ Association rule
2. Classification มาจำแนกกลุ่ม ให้ข้อมูลแต่ละอันเข้า class ใดกับข้อมูล
3. Cluster Analysis แบ่งกลุ่ม