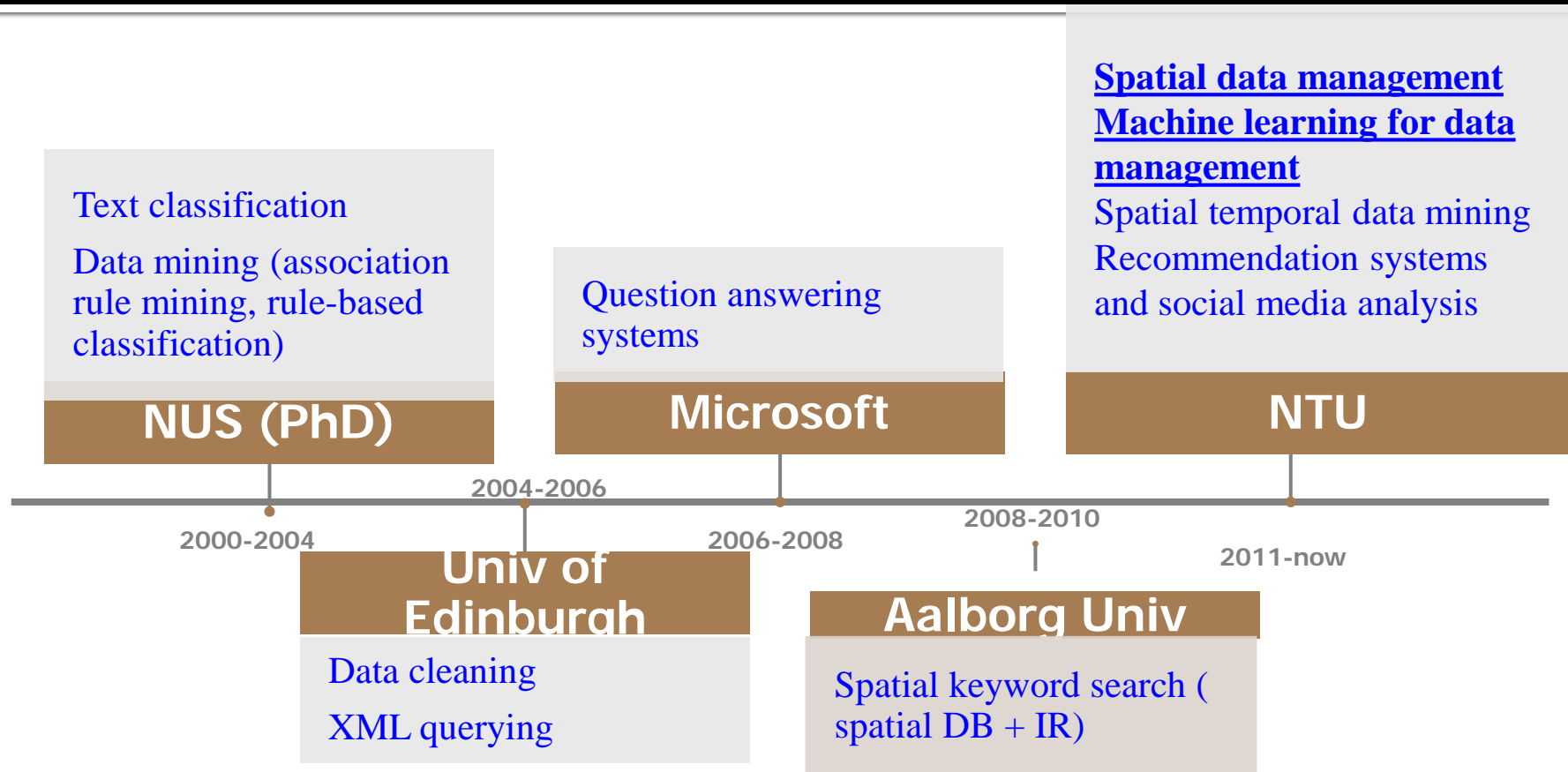


Introduction to second half of SC₄₀₂₀-CE₄₀₃₂-CZ₄₀₃₂ Data Analytics and Mining

About me

- Instructor: Prof Cong Gao
- Specialized in data science (data management, data mining)
- Email: gaocong@ntu.edu.sg
- Homepage:
<https://personal.ntu.edu.sg/gaocong/>
- Office: N4-2c-103
- **Emails: Pls put [SC4020] as email title**
- **You are encouraged to use Discussion Board in the course website so that others can read/answer your questions as well**

My Research Areas



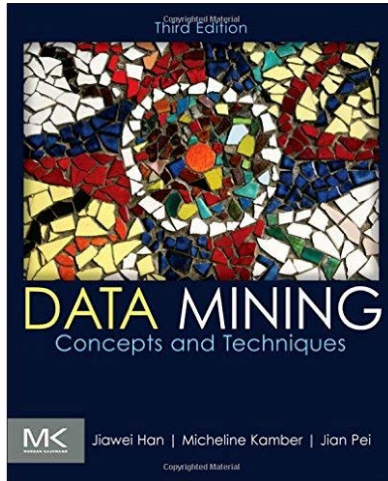
Topics of second half/ tutorial

- Lecture:
 - Association Rule Mining (week 7—8)
 - Sequential pattern mining, Classification, and overfitting (week 9--10)
 - Recommendation Systems (Week 11)
 - Data Processing, Data Cleaning and integration (Week 12)
 - Advanced topic (Week 12)
- **Tutorial will start from Week 7. No tutorial in Week 11**

Assessment (some content is from first lecture)

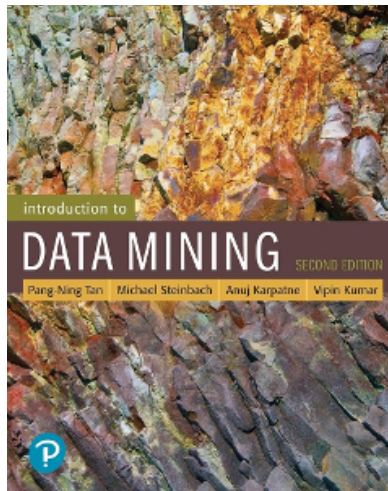
- Final exam (50%) + Assignments (50%)
 - Closed book exam
- Group-based Assignment for 2nd half: 25%
 - Will release toward the end of Week 8
 - Due before the start of final exam (21 Nov, Thursday)
 - Individual Contribution Claim (agreed by all)
 - If there is a large difference, get different scores
 - Plagiarism → 0 marks.

Two Additional Reference Books



**Data Mining: Concepts and Techniques (3rd ed),
Morgan Kaufmann, 2011**

- **Jiawei Han, Micheline Kamber and Jian Pei**



Introduction to Data Mining (Second Edition)

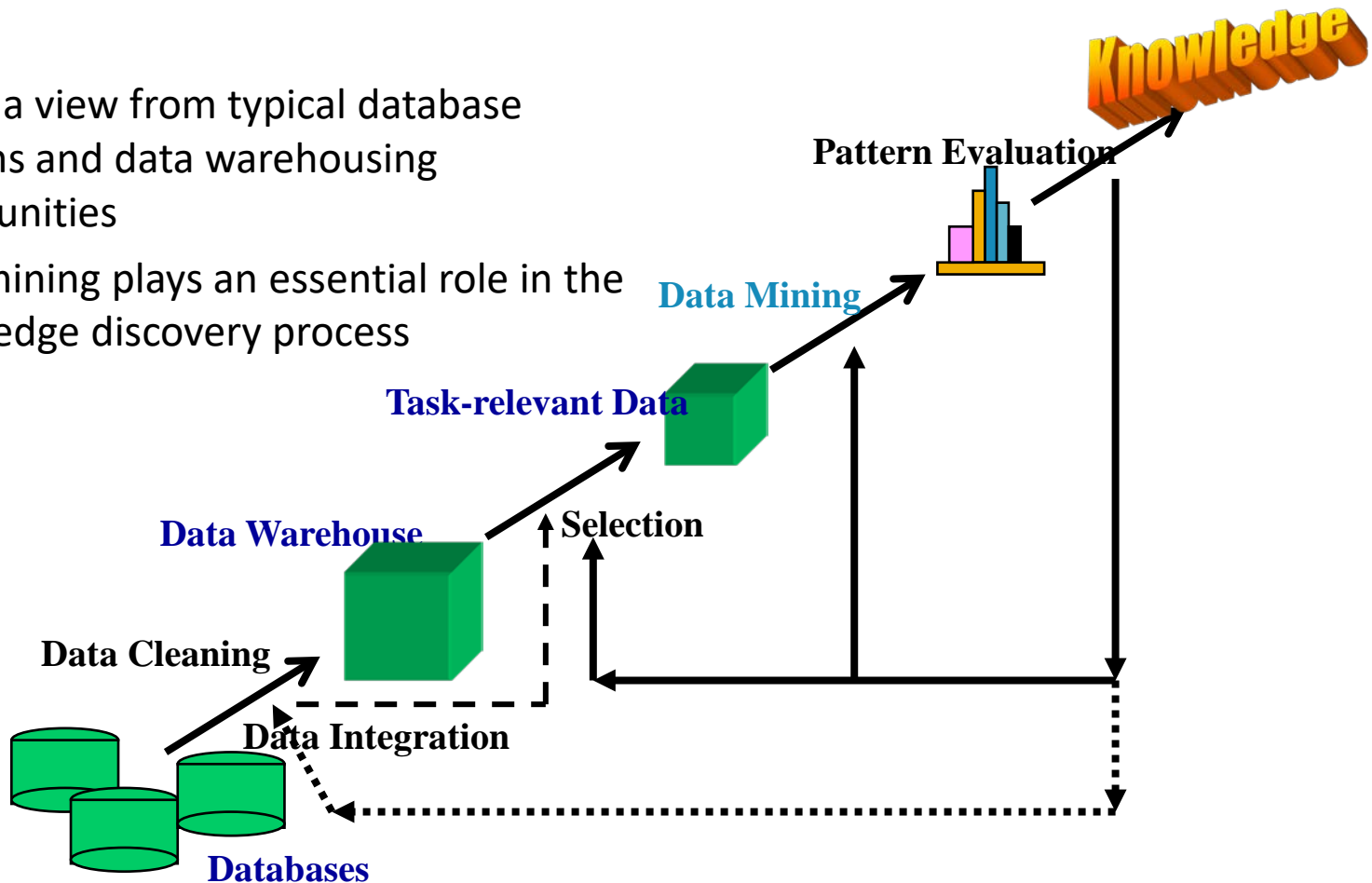
- **by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar**

Additional slides

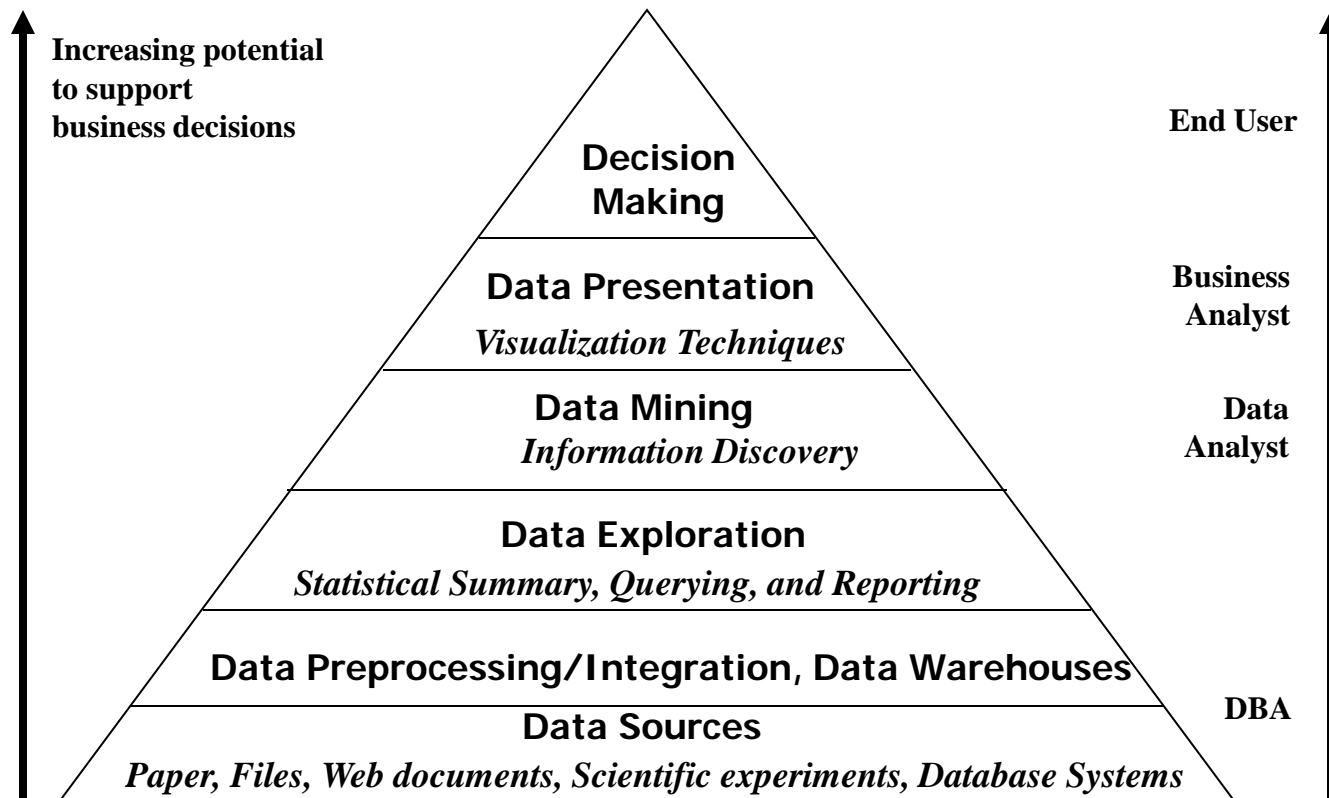
- The rest slides are just for your information

Knowledge Discovery (KDD) Process

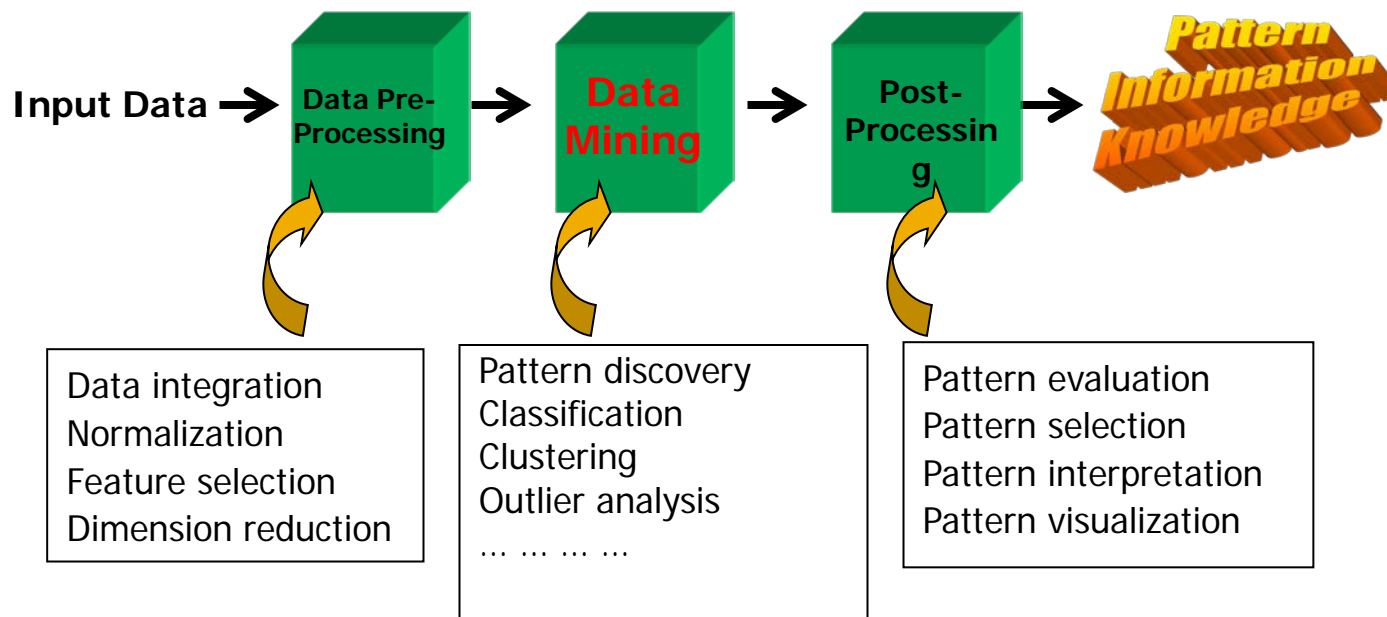
- This is a view from typical database systems and data warehousing communities
- Data mining plays an essential role in the knowledge discovery process



Data Mining in Business Intelligence



KDD Process: A View from ML and Statistics



- This is a view from typical machine learning and statistics communities

A Brief History of Data Mining Society (for your reference)

- 1989 IJCAI Workshop on Knowledge Discovery in Databases
 - Knowledge Discovery in Databases (G. Piatetsky-Shapiro and W. Frawley, 1991)
- 1991-1994 Workshops on Knowledge Discovery in Databases
 - Advances in Knowledge Discovery and Data Mining (U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 1996)
- 1995-1998 International Conferences on Knowledge Discovery in Databases and Data Mining (KDD'95-98)
 - Journal of Data Mining and Knowledge Discovery (1997)
- ACM SIGKDD conferences since 1998 and SIGKDD Explorations
- More conferences on data mining
 - PAKDD (1997), PKDD (1997), SIAM-Data Mining (2001), (IEEE) ICDM (2001), WSDM (2008), etc.
- ACM Transactions on KDD (2007)