

大数据分析

课程内容设置

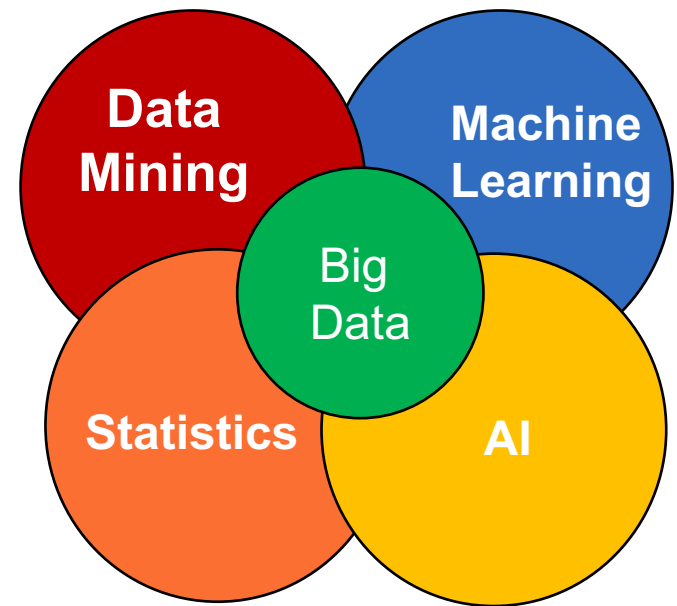
程学旗

靳小龙

刘盛华

课程内容

- This class overlaps with data mining, machine learning, statistics, artificial intelligence but more stress on
 - **Scalability** (big data)
 - **Algorithms**
 - **Computing architectures**
 - **Automation for handling large data**



课程内容

- Different cultures:
 - To a DB person, it is an extreme form of **analytic processing** – queries that examine large amounts of data
 - Result is the query answer
 - To a ML person, it is the **inference of models**
 - Result is the parameters of the model
- In this class we will do both!

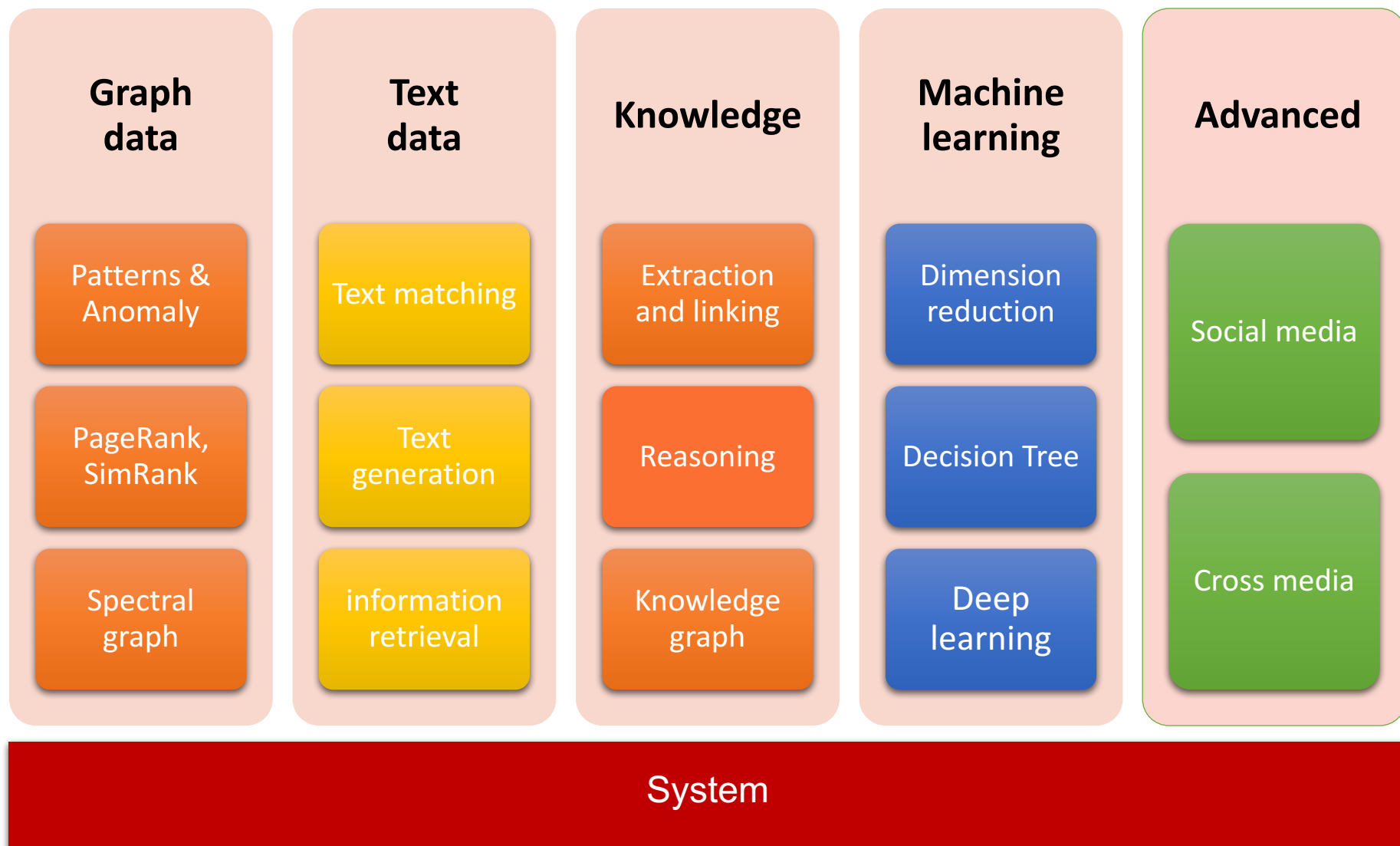
我们将学到什么

- We will learn to **analyze different types of data:**
 - Data is a graph
 - Data is text
 - Data maps to knowledge
- We will learn to **use different models of computation:**
 - MapReduce
 - Streams and online algorithms
 - Single machine in-memory

我们将学到什么

- We will learn **some advanced topics (mixed types of data)**:
 - Social media analytics
 - Cross media analytics
- We will learn **various “tools”** :
 - Linear algebra (SVD, Random walk, Spectral graph)
 - Optimization (stochastic gradient descent, heuristics)
 - Knowledge graphs

内容分布



课程排期

	时间	内容
1	0903	数据科学与大数据概论 大规模数据计算简介
2	0910	大规模机器学习
3	0917	大图挖掘 I
4	0924	大图挖掘 II 流式数据、时间序列分析
	国庆放假	
5	1008	文本数据处理
6	1015	知识计算 I
7	1022	知识计算 II
8	1029	期中复习和评估
9	1105	社交媒体分析
10	1112	跨媒体数据分析
11	1119	大数据分析系统
12	1126	大数据分析应用、总结与展望
13	1203	大作业：分组Poster报告、Q&A
14	1210	闭卷考试