授業科目名 <英訳>	Distributed Information Systems Distributed Information Systems	担当者所属・職名・氏名	情報学研究科 教授　吉川　正俊 情報学研究科 准教授　馬　強

配当学年	1回生以上	単位数	2	開講年度・開講期	2017・後期	曜時限	水3	授業形態	講義	使用言語	英語

授業種別	専攻専門科目

[授業の概要・目的]

This course gives an overview of three major topics on distributed information systems. The first topic is complex data, XML and RDF. Unlike flat tables employed by relational databases, modern information systems manages complex data. Students will learn data models which have rich expressive power to model complex data, and declarative languages to manage complex data. XML (Extensible Markup Language), a W3C standard meta-language for information exchange on the Web and RDF are covered. The second topic covers highly-scalable distributed file systems and databases. The systems covered in lectures include HDFS, MapReduce and Dremel. Column store technologies are also covered as an important storage model for handling OLAP tasks on high-volume data. The third topic is Web mining and knowledge discovery. The fundamental technologies and application systems will be introduced. Some other contemporary topics are lectured if time allows.

[到達目標]

Our goal is to introduce students to principles and techniques of distributed information systems. Students are expected to obtain fundamental knowledge on representation, management, processing and mining of large amount of distributed data.

[授業計画と内容]

Distributed and Parallel Information Systems (10 Lectures by Yoshikawa)
Complex Data, XML and RDF
. Nested Data, Complex Value, Semi-Structured Data
. XML
. RDF
Highly-Scalable Distributed File Systems and Databases
. HDFS (Hadoop Distributed File System)
. MapReduce
. Column Store
. Dremel

Knowledge Discovery (Web Mining) (5 Lectures by Ma)
. Content Mining: Information Extraction, Information Integration (Schema Matching)
. Structure Mining: Link analysis, Social Network Analysis
. Usage Mining: log analysis, personalization, user behavior analysis, HCI
. Sentiment Analysis and Opinion Mining
. Application Systems

Distributed Information Systems(2)へ続く↓↓↓

Distributed Information Systems(2)

[履修要件]

Basic knowledge about database systems.

[成績評価の方法・観点及び達成度]

Grading method: Grade is evaluated by writing examination and reports.

[教科書]

Lecture notes and related documents

[参考書等]

（参考書）

[授業外学習（予習・復習）等]

In some lectures, homework is assigned. Course review is highly recommended.

（その他（オフィスアワー等））

Contact by e-mail using the following addresses:
(Replace AT by @)

Masatoshi Yoshikawa <yoshikawaATi.kyoto-u.ac.jp>
Qiang Ma <qiangATi.kyoto-u.ac.jp>
Toshiyuki Shimizu <tshimizuATi.kyoto-u.ac.jp>

※オフィスアワーの詳細については、KULASIS**で確認してください。**