

Search and Data Mining: Techniques

About

Anna Yarygina

Boris Novikov

Objectives

- Provide a "Big picture" of data analytics
- Overview of techniques and algorithms for analytical problems
- Hands-on approach to different analytical tools
- Develop skills in critical comparative evaluation of analysis results
- Make it easy for those who work during the entire semester
- Make it hard for those who don't.

Организация курса

- Теоретические занятия (в основном лекции)
 - Обсуждение методов, подходов, алгоритмов
 - Доклады студентов на теоретических занятиях (далее «Большие доклады, 45 минут)
- Практические занятия
 - Обсуждение инструментов для выполнения проектов
 - Индивидуальные задания для проектов
 - Проверка результатов выполнения проектов
 - Короткие доклады (15 мин.)
 - Проекты, выполняемые самостоятельно

Scores and Credits

Items	Max. Score
Projects (each of 5)	20
Long presentation (45 min)	15
Short presentation (15 min)	5
Final test (exam)	20

Final Scores	
Excellent	> 81
Good	61 - 80
Satisfactory	40 – 60

How to reduce the scores	
Low quality of the project report	-5
Partially incomplete project	-10
Late submission of a project (more than 2 weeks from starting point)	-5
Late submission of a project (more than 4 weeks from starting point)	-10

Potential Long Presentations

- OLAP
- Association rules
- Search and retrieval
- Text processing
- Classification
- Data streams
- Clustering
- Applications

Projects

Project #	Topic	Start Date	Due date
1	Data pre-processing	Feb. 21	Mar. 06
2	OLAP	Mar 07	Mar 20
3	Text search	Mar 21	Apr 03
4	Classification	Apr 04	Apr 17
5	Clustering	Apr 18	May 03

To submit a completed project, e-mail report together with scripts to both A and B (use CC: in the same message).

It is a student responsibility to compensate for network and other technical failures.

Reading

- Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques, 3rd Edition. The Morgan Kaufmann Series in Data Management Systems, 2012
- Gordon S. Linoff, Michael J. A. Berry -Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3rd Edition. Wiley Publishing, Inc, 2011
- Jesse Russell. Evolutionary data mining. Book on Demand, 2012
- Арменак Барсегян, Михаил Куприянов, Валентин Степаненко, Иван Холод. Технологии анализа данных. Data Mining, Text Mining, Visual Mining, OLAP. БХВ-Петербург. 2007
- Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце. Введение в информационный поиск. Вильямс. 2010. Пер. с англ. Д. Ключина.
- David A. Grossman, Ophir Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 1998