Правительство Российской Федерации

Федеральное государственное автономное учреждение высшего профессионального образования

«Национальный исследовательский университет «Высшая школа экономики»

Факультет Бизнес-информатика

Программа дисциплины

" Технологии извлечения знаний из большого объема данных "/   
“ Knowledge Discovery in Data at Scale Technologies ”

( - на английском языке -)

для направления 38.04.05. Бизнес-информатика

магистерская программа «Системы больших данных»/  
”Big Data systems”

Author of the Program:

Assoc. Prof., PhD. Zinaida K. Avdeeva

[avdeeva@hse.ru](mailto:avdeeva@hse.ru)

Рекомендована секцией УМС Одобрена на заседании кафедры

по бизнес-информатике Инноваций и бизнеса в сфере ИТ

Председатель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Зав. кафедрой С.В. Мальцева

« » 2014 г. « » 2014 г.

Утверждена Ученым Советом факультета Бизнес-информатики

Ученый секретарь \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
« » 2014 г.

Москва - 2014

# **I. Summary**

# General Information about Training Course:

The syllabus is prepared for teachers responsible for the course (or closely related disciplines), teaching assistants, students enrolled on the course as well as experts and statutory bodies carrying out assigned or regular accreditations in accordance with

• educational standards of the State educational budget institution of the Higher Professional Education "State University – The Higher School of Economics" (HSE) categorized as "National Research University",

• Big Data systems curriculum ("Business Informatics", area code 38.04.05.) 1rd year, 2014-2015 academic year;

• Federal state educational standard of higher education in software engineering (Bachelor of Science degree) approved by the Ministry of Education and Science of RF (Russian Federation) Directive N544 on November 9th, 2009 (in Russian).

The course is offered to students of the Master Program «Big Data Systems» (code) in the Faculty of Business Informatics of the National Research University - Higher School of Economics (HSE).

It is a two module course, which is delivered in modules #1 and #2 of the first academic year. Number of credits is 4. Total course length is **144** academic hours including **48** auditory hours (**24** Lecture (L) hours and 24 Practice (P) hours) and **96** Self-study (**SS**) hours. Academic control forms are two home assignment, one test after 1st module , and one written exam after 2nd module.

# Prerequisites:

The course is based on the knowledge of foundations of principles and skills:

* Basic computer science principles and skills,
* Basics in data analysis.

# Interrelation with other disciplines

* Advanced Data Analysis&Big Data for Business Intelligence
* Natural Language Processing.

# Course Objectives and tasks:

# The creation of knowledge from structured and unstructured sources, knowledge discovery is an important discipline of Enterprise Knowledge Management that is focuses on issues of organizational transformation, change and managing knowledge within organizations. The transformation of large amounts of data into valuable information assets directly linked with the technology of knowledge extraction.

# It covers the use of natural knowledge processes of the organization in transforming the organization into a learning organization. It will systematically consider each of the stages of the KM cycle (vision and search, generation, acquisition, capture, transformation, transfer, application) and assess how they relate to the organizational performance.

**Learning outcomes and competences:**

# Competences

|  |  |
| --- | --- |
| Knowledges Skills | Upon successful completion of this course, students will be able to:-udestand role of knowledge in business value creation and main principles knowledge manadgement for strategy business control,- know variety of KM systems and KM circle,- understand principles of choosing IT-technologies to KM support,- understand the role of KD for data analysis,- understand the basic principles of various data mining algorithms,- understand the basic methods of evaluation of created models,- understand basic preprocessing operations,- formulate and solve KD tasks for real-world data,- use selected data mining systems. |
| Practice | To be able to apply the essense tools and techniques of KD in practice. |

**II. Topic-wise Curricula Plan**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Title of the topic | Total Hours | Classroom hours | | Self-study |
| Lectures | Practice |
| Module #3 | | | | | |
| **1** | Introduction to Knowledge Management | 4 | 2 | 2 | 11 |
| **2** | Technologies for knowledge management | 4 | 2 | 2 | 11 |
| **3** | An Architecture for Knowledge Discovery | 6 | 3 | 3 | 13 |
| **4** | Knowledge Discovery methods | 6 | 3 | 3 | 13 |
| **Module #3 totals** | | 24 | **10** | **10** | 46 |
| Module #4 | | | | | |
| 5 | Knowledge Discovery methods | 20 | 6 | 5 | 12 |
| 6 | Special issue | 8 | 7 | 4 | 4 |
| 7 | Technological framework for knowledge manedgement in different types of systems | 8 | 1 | 1 | 4 |
| **Module #4 totals** | |  | **14** | **14** | **50** |
|  | Total: | 144 | 24 | 24 | 96 |

**III. Base Textbook(s) and Recommended Readings**

1. Kenneth E. Kendall, Julie E. Kendall. Systems Analysis and Design. Prentice Hall, 8th ed., 2010.
2. Lesson M. System Analysis and Design. Chicago: Science Research Associates, 1981.
3. Gane C., Sarson T. Structured Systems Analysis and Design. New York: Improved Systems Technologies, Inc., 1977.
4. Naomi Stanford. Organization Design: Collaborative Approach. Elsiever, 2005. P 322
5. Naomi Stanford. Guide to Organization Design: Creating high-performing and adaptable enterprises (The Economist). Bloomberg Press; 1 edition, 2007
6. Harry J. Rosenblatt. Systems Analysis and Design (with CourseMate Printed Access Card) (Shelly Cashman). Cengage Learning, 10 edition, 2013.
7. Gareth R. Jones. Organizational Theory, Design, and Change, 6/E, Publisher: Prentice Hall, Copyright: 2010.

**IV. Forms of Control**

Current control: attendance record, seminar-based knowledge control, group project control;

- Intermediate control: written test by the end of Module 1, group project 1 presentation by the end of Module 2;

- Final control: exam by the end of Module 2;

- The final course grade is a sum of the following elements:

1) attendance record (A);

2) practice activities (S);

3) group projects (P1, P2);

4) written test (T);

5) exam (E).

The overall and accumulated course grades Go and Ga (10-point scale each) are calculated as follows:

Ga = 0.15A + 0.15S + 0.3(P1+P2) + 0.4T;

Go = 0.5Ga + 0.5E.

The overall and accumulated course grades Go and Ga (10-point scale each) include results achieved by students in their attendance record A, practice activities S, group project P, written test T and exam E; it is rounded up to an integer number of points. The rounding procedure accounts for students' practice activities during seminars. Intermediate assessment retakes are not allowed. Conversion of the concluding rounded grade (FE) to five-point scale grade is done in accordance with the following table:

Summary Table : Correspondence of ten-point (10) to five-point (5) system's marks

|  |  |
| --- | --- |
| Ten-point scale [10] | Five-point scale [5] |
| 1 - unsatisfactory  2 - very bad  3 - bad | Unsatisfactory (UnS) - 2 |
| 4 - satisfactory  5 - quite satisfactory | Satisfactory (S) - 3 |
| 6 - good  7 - very good | Good (G) - 4 |
| 8 - nearly excellent  9 - excellent  10 - brilliantly | Excellent (E) - 5 |

**V. Course Contents**

* + - Тopic 1: **Introduction to Knowledge Management.**

♦ Outline:

* Course: Structure, the basic goals and outcomes, competences profile of the system analyst. The main area of knowledge founding the course subjects, active theorist, methodologists and practices.

• Data, information and knowledge

• Types of knowledge

• Knowledge Management Models

• Organizational Memory

• Knowledge creation and organizational learning

• Knowledge codification

• Knowledge sharing and transfer

• Business and competitive intelligence

• Ethical issues in knowledge management

• The Role of Culture in Knowledge Management

• Business Process and Knowledge Management

• Alignment of Business and Knowledge Management Strategies

• Intellectual capital and knowledge management

• Measurement of impact of knowledge management programs

♦ Core books (sources of information)

* 1. Dalkir K. Knowledge management in theory and practice - 2th. ed. — London: Massachusetts Institute of Technology, 2011. — 486 p. — ISBN 9780262015080
     + Тopic 2: **Technologies for knowledge management**:

♦ Outline:

• Artificial Intelligence, Digital Libraries, Repositories, ECM, etc

Knowledge-Based Systems

• Information/knowledge audit

• Data mining. Ontology-supported web service composition

•

♦ Core books (sources of information)

1. Dalkir K. Knowledge management in theory and practice - 2th. ed. — London: Massachusetts Institute of Technology, 2011. — 486 p. — ISBN 9780262015080

Тopic 3: **An Architecture for Knowledge Discovery**

♦ Outline:

* KM cycle (vision and search, generation, acquisition, capture, transformation, transfer, application).
* Knowledge Capture Systems: Systems that Preserve and Formalize Knowledge; Concept Maps, Process Modeling, RSS, Wikis, Delphi Method, etc
* Knowledge Sharing Systems: Systems that Organize and Distribute Knowledge; Ontology Development Systems, Categorization and Classification Tools, XML-Based Tools, etc
* Data mining. Ontology-supported web service composition
* Data analisys

♦ Core books (sources of information)

1. Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities (Studies in Big Data) Wesley W. Chu.(editor), Springer. 2013

* 1. The Data mining and knowledge discovery handbook, 2nd ed - Oded Maimon; Lior Rokach, editors, Springer, 2010

 Тopic 4: **Knowledge Discovery methods.**

♦ Outline:

* Types of Discovery: correlation, class, novelty, association
* Preprocessing Methods
* Supervised Methods
* Unsupervised Methods
* Soft Computing Methods
* Supporting Methods
* Advanced Methods

♦ Core books (sources of information)

* 1. Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities (Studies in Big Data) Wesley W. Chu.(editor), Springer. 2013
* The Data mining and knowledge discovery handbook, 2nd ed - Oded Maimon; Lior Rokach, editors, Springer, 2010
  + - Тopic 5**. Special issue**.

♦ Outline:

* Aspect and entity extraction for opinion mining
* Mining periodicity from dynamic and incomplete spatiotemporal data
* Mining discriminative subgraph patterns from structural data
* A social search engine
* An approach for Social network Integration and Analysis with Privacy Preservation
* A clustering approach to constrained binary matrix factorization
* .

♦ Core books (sources of information)

1. Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities (Studies in Big Data) Wesley W. Chu.(editor), Springer. 2013

The Data mining and knowledge discovery handbook, 2nd ed - Oded Maimon; Lior Rokach, editors, Springer, 2010

**V. Topics for course final quality assessment**

1) Exam topics:

1. Data, information and knowledge
2. Types of knowledge
3. Knowledge Management Models
4. Organizational Memory
5. Knowledge creation and organizational learning
6. Knowledge codification
7. Knowledge sharing and transfer
8. Business and competitive intelligence
9. Ethical issues in knowledge management
10. The Role of Culture in Knowledge Management
11. Business Process and Knowledge Management
12. Alignment of Business and Knowledge Management Strategies
13. Intellectual capital and knowledge management
14. Technologies for knowledge management: basic concept
15. Artificial Intelligence, Digital Libraries, Repositories, ECM, etc
16. Knowledge-Based Systems
17. Information/knowledge audit
18. An Architecture for Knowledge Discovery: basic concept
19. KM cycle (vision and search, generation, acquisition, capture, transformation, transfer, application).
20. Knowledge Capture Systems: Systems that Preserve and Formalize Knowledge; Concept Maps, Process Modeling, RSS, Wikis, Delphi Method, etc
21. Knowledge Sharing Systems: Systems that Organize and Distribute Knowledge; Ontology Development Systems, Categorization and Classification Tools, XML-Based Tools, etc
22. Data mining. Ontology-supported web service composition
23. Data analisys
24. Knowledge Discovery methods: basic cocepts, types
25. Types of Discovery: correlation, class, novelty, association
26. Preprocessing Methods
27. Supervised Methods
28. Unsupervised Methods
29. Soft Computing Methods
30. Supporting Methods
31. Advanced Methods

**V. Training materials**

* 1. Glosary.
* 2. Guidelines for the implementation of practical tasks.
* 3. Structure of the report for the tasks with forms to fill results.
* **VI. Assignment topics**
* Home assignment for group project:
* The course includes two group projects (one in each year of study), compulsory to all students. Students will work in groups of up to 4 students on one of the suggested topics.
* First project: Research business cases : understanding the peculiarity of the corresponding organizational system and requirements of the supporting information system .
* Second project: Caring through 5 phase of design .
* Written tests:
* The written tests is a testing assignment based on the covered topics.

The author of program: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Z.K. Avdeeva