|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Al-Farabi Kazakh National Univesity**  **Syllabus**  **MCAO5207** **Semantic-based analysis for Information Security**  **Spring semester 2018-2019 academic year**  **1 course, Master degree, spec. Information systems security** | | | | | | | | | | |
| The code of the course | The name of the course | Type | The number of hours per week | | | | | | The number of credits | ECTS |
| Lecture | | Practice | | Laboratory | |  | MCAO5207 |
| MCAO5207 | Semantic-based analysis for Information Security | BC | 2 | | 1 | | 0 | | 3 | 6 |
| Prerequisites | algorithms and data structure, databases | | | | | | | | | |
| Laboratory work | Vladislav Karyukin | Office hours | Monday – 18:00 – 18:50 | | | | | | | |
| Email | vladislav.karyukin@gmail.com |
| The mobile phone | 8 701 940 59 92 | Room | 324 | | | | | | | |
| Academic presentation of the course | **A type of the academic course** (practical; basic) this course allows students to get knowledge of word’s morphology models, context-free grammars and n-gram model of word prediction, models and methods part of speech tagging and syntax parsing.  **The purpose of the course**: is to provide theunderstanding of the principles and practices of formal grammar of natural languagesincluding a finite-state transducers model of word’s morphology, context-free grammars and n-gram model of word prediction, models and methods part of speech tagging and syntax parsing, semantics and unification grammars, models and methods of computational lexical semantics.  This course forms the following competences in a context of a qualified specialist:  А) cognitive: be able  - to demonstrate received knowledge (lexicon, morphotactic, morphological parsing, finite-state automata, probabilistic models, N-grams, part of speech tagging, context-free grammar)  - to demonstrate understanding the whole structure in a field of study and be able to find connections between its elements (analysis of words in isolation and in a context)  Б) functional: be able  - include new knowledge in a context of basic knowledge of a specialist and interpret its content;  - analyze academic situation and offer directions of its solutions;  - use methods (research, analysis, calculations and so on) that are specific to a field of study (creating words with morphemes, syntactic parsing large sentences, finding probabilities of specific sentences in a large corpus of texts) in an individual or group study activities  В) system: be able  - Generalize, interpret and evaluate the learning outcomes obtained in the context of the discipline, the training module, the content of midterm exam (to analyze and create models for concrete part of natural language, to create algorithms of semantic analysis of natural language; to have an idea: to improve formal models of natural languages);  - analyze dynamic of scientific problems’ solutions (scientific overviews of specific problems’ research)  - Make an analysis of the results of the course, summarize them in the form of a scientific essay, presentation, review, scientific review and so on;  Г) social: be able  - to constructive academic and social communications and collaborations in groups; propose a problem and tell why it is important;  - teamwork  Д) metacompetence: be able  - To realize the role of the listened course in the implementation of the individual trajectory of education. | | | | | | | | | |
| Prerequisites | Information and communication technologies | | | | | | | | | |
| Literature and resources | **Basic:**  Jurafsky D., Martin J. Speech and language processing: an introduction to naturel language processing, computational linguistics, and speech recognition. Pearson, Prentice hall.2009, 988p.  **Additional:**   1. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. - М.: Изд. дом "Академия". 2006.- 304 с. 2. Минский М. Фреймы для представления знаний. – М.: Энергия. 1979. – 150 с. 3. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 1: Синтаксический анализ.М.: Мир, 1978. — 612 с. 4. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 2: Компиляция. М.: Мир, 1978. — 487 с. | | | | | | | | | |
| The academic policy of the course in the context of university moral and ethical values | **Rules of academic behavior**  Mandatory presence in the classroom, inadmissibility of late arrivals. Absence and delay in classes without prior warning of the teacher are estimated at 0 points. Mandatory compliance with the deadlines for the completion and delivery of assignments (for CDS, boundary, control, laboratory, project, etc.), projects, examinations. In case of violation of the deadlines, the task is evaluated taking into account the deduction of penalty points.  **Academic value**:  Academic honesty and integrity: doing all tasks independently; Inadmissibility of plagiarism, forgery, cheating at all stages of knowledge control, deception of the teacher and disrespectful attitude towards him. (Code of Honor for a student of KazNU) | | | | | | | | | |
| The course requirements | 1. The assignments will be distributed throughout the whole semester as it is shown in the schedule of the course 2. Assignments are given individually for every student. They contain 30% of the final grade 3. You will use the study material during the whole semester in practical purposes 4. Assignments have to be completed in time 5. Assignments are implemented in C++ programming language, and they are shown for evaluation during laboratory classes 6. The course work will contain 10% and 25% of the final grade | | | | | | | | | |
| The grade policy | Description of works | | | Weight | | | | | | |
| Assignments  Laboratory work  Exams  Aggregate | | | 33%  27%  40%  100% | | | | | | |
|  | Your final grade will be calculated using this formula  The final grade of the course =  There are grades in percentages:  A: 95%-100%  A-: 90%-95%  B+: 85%-89%  B: 80%-84%  B-: 75%-79%  C+: 70%-74%  C: 65%-69%  C-: 60%-64%  D+: 55%-59%  D: 50%-54%  F: 0%-49% | | | | | | | | | |
| The course policy | Corresponding dates of laboratory works and assignments can be prolonged in cases of extenuating circumstances (illness, emergency cases, accidents, unseen circumstances) according to the academic policy of the university. | | | | | | | | | |
| **The schedule of the course (Application 1)** | | | | | | | | | | |
| Week | A name of the topic | | | | | A number of hours | | Maximum points | | |
| 1 | **Lecture(L) 1.** Introduction, morphology and regular expressions | | | | | 2 | | 4 | | |
|  | **Laboratory work (LW) 1.** Morphology and regular expressions  Examples. | | | | | 1 | | 6 | | |
| 2 | **L2.** Finite-state transducers(FST). FST for morphological parsing. | | | | | 2 | | 4 | | |
|  | **LW2.** morphological parsing, finite-state automata) | | | | | 1 | | 6 | | |
| 3 | **L3.** Finite-state transducers and orthographical rules. Lexicon-free FST: The Porter stemmer. | | | | | 2 | | 4 | | |
|  | **LW3.** Word errors and probabilities | | | | | 1 | | 6 | | |
|  | **MSWT**1. Morphology and regular expressions Finite-state morphological parsing, Finite-state transducers | | | | | 1 | | 12 | | |
| 4 | **L4**. Word counting in corpora. Simple n-grams. Training and test sets. | | | | | 2 | | 4 | | |
|  | **LW4**. N-grams | | | | | 1 | | 6 | | |
| 5 | **L5**. Smoothing models and methods | | | | | 2 | | 4 | | |
|  | **LW5**. Smoothing models | | | | | 1 | | 6 | | |
|  | **MSWT2**. Word counting in corpora, smoothing models. | | | | | 1 | | 8 | | |
| 6 | **L6**. English word classes. Tagsets of English. Part-of-speech tagging. Rule-based part-of-speech tagging. | | | | | 2 | | 4 | | |
|  | **LW6**. Tags | | | | | 1 | | 6 | | |
| 7 | **L7**. Context-free grammars. Some grammar rules for English. | | | | | 2 | | 4 | | |
|  | **LW7**. Context-free parsing | | | | | 1 | | 6 | | |
|  | **MSWT3**. Stochastic part of speech tagging. | | | | | 1 | | 10 | | |
|  | **1 Boundary control** | | | | |  | | 100 (COEF. 0.3) | | |
|  | **Midterm exam** | | | | |  | | 100 (COEF. 0.1) | | |
| 8 | **L8**. Syntax parsing. Partial parsing. | | | | | 2 | | 4 | | |
|  | **LW8**. Top-down and bottom-up approaches in syntax parsing | | | | | 1 | | 6 | | |
| 9 | **L9**. HMM part-of-speech tagging. Transformation-based tagging | | | | | 2 | | 4 | | |
|  | **LW9**. HMM part-of-speech tagging | | | | | 1 | | 6 | | |
|  | **MSWT4**. Rule-based part-of-speech tagging. Transformation tagging | | | | | 1 | | 8 | | |
| 10 | **L10**. Basics of model - theoretical semantics. Logic of the first order, as the tools of representation of meaning | | | | | 2 | | 4 | | |
|  | **LW10**. Logic of the first order | | | | | 1 | | 6 | | |
| 11 | **L11**. Event and state representations. | | | | | 2 | | 4 | | |
|  | **LW11**. Events and state representation | | | | | 1 | | 6 | | |
|  | **MSWT5**. Basics of model - theoretical semantics | | | | | 1 | | 8 | | |
| 12 | **L12**. Features, unification method and unification grammars. Syntax-driven semantic analysis. Semantic augmentations to syntactic rules. | | | | | 2 | | 4 | | |
|  | **LW12**. Semantic augmentations to syntactic rules. Examples. | | | | | 1 | | 6 | | |
| 13 | **L13**. Word sense disambiguation(WSD). Supervised WSD. | | | | | 2 | | 4 | | |
|  | **LW13**. Supervised WSD | | | | | 1 | | 6 | | |
|  | **MSWT6**. Features, unification method and unification grammars. Syntax-driven semantic analysis. | | | | | 1 | | 8 | | |
| 14 | **L14**. WSD: dictionary and thesaurus methods. WSD: bootstrapping. | | | | | 2 | | 4 | | |
|  | **LW14**. WSD: bootstrapping | | | | | 1 | | 6 | | |
| 15 | **L15**. Word similarity: thesaurus methods. Word similarity: distributional methods. | | | | | 2 | | 2 | | |
|  | **LW15**. Word similarity: distributional methods. | | | | | 1 | | 2 | | |
|  | **MSWT7**. Word similarity: thesaurus methods. Word similarity: distributional methods. | | | | | 1 | | 2 | | |
|  | 2 Boundary control | | | | |  | | 100 (COEF. 0.3) | | |
|  | Exam | | | | |  | | 100 (COEF. 0.4) | | |
|  | Totally | | | | |  | | 100 | | |

**Application 2**

**Description of expected results as a system of formed competence**

**(Dublin descriptors)**

**Cognitive competence:**

**Knowledge:** A) Memorization and reproduction of the studied material - from concrete facts to integral theory, B) organize, define, repeat, fill tables, memorize, name, order, recognize, treat, remind, repeat, reproduce, make a list, highlight, tell, show and so on.

**Perception:** A) an ability to convert materials from one form of expression to another one, interpret information, make an assumption about a further course of events; B) Classify, describe, discuss, explain, express, certify, find, recognize, report, reformulate, analyze, choose, translate

**Functional competence:**

**Application**. A) an ability to use the material studied in specific conditions and new situations; B) apply, select, demonstrate, dramatize, illustrate, interpret, work, practice, develop a schedule, decide, use, write.  
**Analysis**. A) an ability to isolate parts of the whole; identify relationships between them; Determine the principles of the organization of the whole; makes a distinction between facts and consequences; B) analyze, evaluate, calculate, classify, compare, criticize, differentiate, distinguish, distinguish, study, experiment, conduct an experiment, identify similarities and differences, clarify parameters, perform a test.

**System competence**:  
**Synthesis**. A) an ability to combine elements in order to obtain a whole with novelty, (composition, speech, report, abstract, project, case, quest, etc.); B) organize, collect, compile, build, create, develop, formulate, prove their point of view, manage, organize, plan, forecast, prepare, offer, create, write.  
**Evaluation**. A) an ability to assess the value of a material, the logic of the presentation of information, the construction of the text, the conformity of conclusions, the significance of a product activity; B) evaluate, discuss, select, compare, protect, predict, select, support, defend a point of view, prove, predict, present the argumentation.

**Social competence**:  
Willingness to cooperate: A) communicate information, ideas, problems and solutions, work in a team; B) formulate (problem, goal, task, conclusions, regulations, etc.); identify (requirements, criteria, principles); Take decisions and report on them, draw conclusions, argue, justify, insist, persuade, etc.

**Metacompetence**:  
Skills in the field of training: A) develop such skills in the field of training, which are necessary for continuing education with a high degree of autonomy; B) Be capable of reflection, an objective assessment of their achievements; To realize the necessity of forming new competencies; Determine the direction of further personal and professional development, etc.

Lector Kulmamirov S. A.

Chair’s chief Yessengalieva Zh. S.

Dean of the faculty Tukeev U. A.