

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
Кафедра теоретичної кібернетики



«ЗАТВЕРДЖУЮ»

Заступник декана з навчальної роботи

Кашпур О.Ф.

«28» серпня 2020 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
КЛАСТЕРИЗАЦІЯ ТА КЛАСИФІКАЦІЯ ІНФОРМАЦІЇ/
CLUSTERING AND CLASSIFICATION OF INFORMATION

для студентів

галузі знань 12 – "Інформаційні технології"/
"Information Technology"
спеціальність 112 – «Комп'ютерні науки»/ Computer Science
освітній рівень магістр/ Master's educational level
освітньо-наукова програма «Математичні Методи Штучного
Інтелекту» / "Mathematical Methods of
Artificial Intelligence"
вид дисципліни обов'язкова/ mandatory

Форма навчання денна
Навчальний рік 2020/2021
Семестр 2
Кількість кредитів ECTS 5

Мова викладання, навчання та оцінювання українська, англійська
Форма заключного контролю іспит/exam

Викладач / Teacher: доцент Трохимчук Р.М., к.ф.-м.н. /
Ph.D., Candidate of physical and mathematical sciences, Associate Professor Trokhymchuk R.M.

Пролонговано: на 2021/2022 н.р. «11» 02 2021 р.
на 20__/20__ н.р. () «__» __ 20__ р.

КИЇВ – 2020

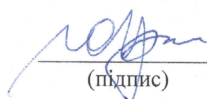
© Трохимчук Р.М., 2020 рік

Розробник:

Трохимчук Ростислав Миколайович, к.ф.-м.н., доцент, доцент кафедри теоретичної кібернетики

ЗАТВЕРДЖЕНО

Зав. кафедри теоретичної кібернетики


(підпис)

(Крак Ю.В.)
(прізвище та ініціали)

Протокол № 4 від «30» Серезня 2020 р.

Схвалено Гарантом освітньо-наукової програми «Математичні методи штучного

інтелекту»


(підпис)

(Завадський І.О.)

«2» квітня 2020 р.

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «2» квітня 2020 року № 1

Голова науково-методичної комісії

(підпис)

(Омельчук Л.Л.)
(прізвище та ініціали)

«2» квітня 2020 року

1. Мета дисципліни: засвоєння теоретичних знань і набуття практичних навичок з сучасної теорії інтелектуального аналізу даних (Data Mining), необхідних всім студентам, що спеціалізуються в галузях прикладної математики та інформатики, математичної кібернетики і в подальшому вивчатимуть такі розділи сучасної інформатики, як системи автоматизованого керування, системи обробки і передачі інформації, аналіз даних, системи штучного інтелекту, робототехніка тощо.

/ **Discipline aim.** Mastering theoretical knowledge and practical skills in the modern theory of data mining (Data Mining), required for all students who specialize in applied mathematics and computer science, mathematical cybernetics and in the future will study such automated sections of computer science, control, information processing and transmission systems, data analysis, artificial intelligence systems, robotics and more.

2. Попередні вимоги до опанування або вибору навчальної дисципліни / Prerequisites for mastering or choosing a course

1. *Знати* базові поняття програмування, теорії баз даних, загальної алгебри, обчислювальної математики, теорії ймовірностей, математичної статистики, аналізу даних. / *Know* basic concepts of programming, database theory, general algebra, computational mathematics, probability theory, mathematical statistics, data analysis.

2. *Уміти* ефективно застосовувати зазначений математичний апарат для розв'язання різноманітних практичних задач. / *Be able to* effectively use the specified mathematical apparatus for solving various practical problems.

3. Анотація навчальної дисципліни / Summary of the course

Предметом навчальної дисципліни «Кластеризація та класифікація інформації» є розгляд основних понять, моделей, результатів і методів сучасної математичної теорії інтелектуального аналізу даних. / **The subject of the course** "Clustering and Classification of Information" is the examination of basic concepts, models, results and methods of modern mathematical theory of Data Mining.

Викладається у 2 семестрі 1 курсу магістратури в обсязі – **150 год.** / It is taught in semester 2 of the master's course of 150 hours.

5 кредитів ECTS, зокрема: *лекції – 42 год., самостійна робота – 106 год., консультацій – 2 год.* / **5 ECTS credits**, in particular: *lectures – 42 hours, independent work – 106 hours, consultations – 2 hours.*

У курсі передбачено одну *контрольну роботу*. / The course provides one test work. Завершується дисципліна – **іспитом** у 2 семестрі 1 курсу магістратури. / Completes the discipline – **exam** in semester 2 of the master's course.

4. Завдання (навчальні цілі):

Набуття знань, умінь та навичок (компетентностей):

- здатність спілкуватися іноземною мовою (ЗК5);
- здатність генерувати нові ідеї (креативність) (ЗК8);
- здатність приймати обґрунтовані рішення (ЗК12);
- здатність ідентифікувати моделі складних систем і процесів, розробляти та застосовувати методи і засоби моделювання та прогнозування систем і процесів в умовах невизначеності (СК2);
- здатність до дослідження та аналізу надвеликих масивів даних із складною неоднорідною і/або невизначеною структурою для прийняття зважених бізнес-рішень (СК3);
- здатність застосовувати методи і засоби організації великих даних для проектування масштабованих інфраструктур консолідації ресурсів зберігання, дослідження, управління, захисту та обслуговування інформації, розв'язання завдань моделювання та прогнозування стратегічних напрямків розвитку бізнесу (СК4);

- здатність вирішувати надскладні наукові та інженерні задачі, що передбачають розпаралелювання обчислень, великих витрат машинного часу, обчислювальних ресурсів і методів організації розв'язання задач на суперкомп'ютерах (СК5).

/ Tasks (learning objectives):

Acquisition of knowledge, skills and competences (competences):

- ability to communicate in a foreign language;
- creativity;
- ability to make reasonable decisions;
- ability to identify models of complex systems and processes, to develop and apply methods and tools for modeling and forecasting systems and processes in conditions of uncertainty;
- ability to research and analyze ultra-large data sets with a complex inhomogeneous and / or uncertain structure for making informed business decisions;
- ability to apply methods and tools of big data organization for design of scalable infrastructures of consolidation of resources of storage, research, management, protection and service of the information, the decision of problems of modeling and forecasting of strategic directions of business development;
- ability to solve extremely complex scientific and engineering problems involving parallelization of computations, high costs of machine time, computing resources and methods of organizing problem solving on supercomputers.

5. Результати навчання за дисципліною / Learning outcomes of the discipline

Результат навчання (РН) (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність) / Learning Outcome (LO) (1. know; 2. be able; 3. communication; 4. autonomy and responsibility)		Форми (та/або методи і технології) викладання і навчання / Forms (and / or methods and technologies) of teaching and learning	Методи оцінювання та пороговий критерій оцінювання (за необхідності) / Assessment methods and assessment threshold (if applicable)	Відсоток у підсумковій оцінці з дисципліни / Percentage of the final grade in the discipline
Код/ Code	Результат навчання/ Learning Outcome			
РН/ LO 1.1	Знати особливості подання, збереження та обробки інформації у сучасних комп'ютерних системах / Know the features of presenting, storing and processing information in modern computer systems	Лекція / Lecture	Контрольна робота (КР), 60% правильних відповідей, icnum / Control, 60% correct answers, exam	20%
РН/ LO 1.2	Знати основні математичні моделі та методи, що застосовуються для обробки та аналізу великих масивів інформації / Know the basic mathematical models and methods used to process and analyze large amounts of information	Лекція / Lecture		
РН/ LO 1.3	Знати технологію проектування ефективних алгоритмів для аналізу інформації / Know the technology of designing effective algorithms for information analysis	Лекція / Lecture	КР, 60% правильних відповідей, icnum / Control,	20%

			60% correct answers, exam	
PH/ LO 1.4	Знати основні структури даних і алгоритми для інтелектуального аналізу даних / Know the basic data structures and algorithms for Data Mining	Лекція / Lecture	Виконання самостійної роботи (CP), іспит / Independent work, exam	10%
PH/ LO 2.1	Вміти вибирати та застосовувати оптимальні в конкретних умовах алгоритми для здійснення кластеризації та класифікації даних / Being able to select and apply the optimal algorithms for specific clustering and classification of data	Лекція, Самостійна робота (CP) / Lecture, Laboratory, independent work	КР, захист CP, іспит / Control, Independent work, exam	20%
PH/ LO 2.2	Вміти здійснювати оцінку ефективності проєктованих алгоритмів кластеризації та класифікації даних / Be able to evaluate the performance of the designed CL clustering and classification algorithms		Захист CP, іспит / Independent work, exam	30%

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання / The correlation of learning outcomes of the discipline with the program learning outcomes

Результати навчання дисципліни / Learning outcomes of the discipline Програмні результати навчання / Program learning outcomes (PLO)	PH/ LO 1.1	PH/ LO 1.2	PH/ LO 1.3	PH/ LO 1.4	PH/ LO 2.1	PH/ LO 2.2
ПРН4. Аналізувати великі дані та моделювати високорівневі абстракції у великих наборах даних різної природи, проєктувати сховища великих даних, для видобутку даних і знань, візуалізовувати великі дані, будувати і оцінювати регресивні моделі, що генеруються на основі великих даних. / Analyze big data and model high-level abstractions in large data sets of different nature, design big data repositories to extract data and knowledge, visualize big data, build and evaluate regression models generated based on big data.	+	+	+	+		
ПРН5. Вирішувати складні проблеми, що вимагають систем з великою обчислювальною потужністю для забезпечення масштабованості паралельних алгоритмів і програм. / Solve complex problems that require systems with high computing power to ensure the scalability of parallel algorithms and programs.		+		+		

ПРН6. Використовувати розподілені високопродуктивні обчислювальні технології для забезпечення ефективного вибору та використання консолідованих ресурсів і послуг. / Use distributed high-performance computing technologies to ensure the efficient selection and use of consolidated resources and services.				+	+	+
ПРН7. Вміти використовувати обчислювальні системи надвеликої потужності для виконання парадигми програмування мультипроцесорних обчислень, розробляти ефективні паралельні алгоритми складних виробничих задач, застосовувати хмарні платформи та їх віртуалізацію. / Be able to use ultra-high-power computing systems to perform the paradigm of programming multiprocessor computing, to develop effective parallel algorithms for complex production tasks, to use cloud platforms and their virtualization.					+	+

7. Схема формування оцінки. / Scheme of formation of assessment.

7.1 Форми оцінювання студентів: Результати навчальної діяльності студентів оцінюються за 100-бальною шкалою.

- семестрове оцінювання:

1. Контрольна робота: РН1.1, РН1.2, РН1.3, РН2.1 – 30 балів / 18 балів,
2. Захист самостійної роботи: РН1.4, РН2.1, РН2.2 – 30 балів / 18 балів.

Підсумковий контроль проводиться у формі іспиту – 40 балів.

Підсумкова оцінка $100 = 60 + 40$.

Якщо студент з поважних причин, які підтверджено документально, був відсутній при написанні контрольної роботи, він має право на одне перескладання з можливістю отримання максимальної кількості балів. Термін перескладання визначається викладачем.

Студент допускається до іспиту, якщо в семестрі набрав не менше ніж 20 балів. Для отримання загальної позитивної оцінки з дисципліни оцінка за іспит має бути не менше 24 балів.

Іспит вважається не зданим, якщо сумарна кількість балів з дисципліни складає менше 60 балів.

На підставі п.п. 4.1, 4.2 Положення про порядок оцінювання знань студентів та п.п. 4.6.1, 7.1.5, 7.1.11, 7.1.12 Положення про організацію освітнього процесу:

- максимальна кількість балів які можуть бути отримані студентом упродовж 1-го семестру: 100 балів;

- студент допускається до іспиту, якщо в семестрі набрав не менше 20 балів та отримав не менше мінімальної порогової кількості балів за поточну та контрольну лабораторну роботу. Для отримання загальної позитивної оцінки з дисципліни оцінка за іспит має бути не менше 24 балів.

– якщо протягом семестру студент набрав менше 60 балів і допущений до підсумкового контролю, то йому надається можливість написати підсумкову контрольну роботу (25/15 балів), бали за яку додаються до набраних упродовж семестру балів;

- форма проведення: письмова робота;

- види завдань: 4 теоретичні питання.

Студент допускається до підсумкового контролю у формі іспиту, якщо в семестрі набрав не менше 20 балів за контрольні та лабораторну роботи і при цьому загалом набрав не менше

24 балів. Для отримання загальної позитивної оцінки з дисципліни має бути отримано не менше 60 балів.

- підсумкове оцінювання: іспит

- максимальна кількість балів які можуть бути отримані студентом: 40 балів;
- результати навчання які будуть оцінюватись: PH1.1, PH1.2, PH1.3, PH1.4, PH2.1, PH2.2;
- форма проведення і види завдань: письмова робота.

Рекомендований мінімум – 24 балів.

/

7.1 Forms of Student Assessment: Students' learning outcomes are rated on a 100-point scale.

1. Test work: LO1.1, LO 1.2, LO 1.3, LO 2.1 - 30 points / 18 points,
2. Defense of independent work: LO 1.4, LO 2.1, LO 2.2– 30 points / 18 points.

The final control is conducted in the form of exam – 40 points.

The total score is $100 = 60 + 40$.

If the student for absent reasons, which is documented, was absent when writing the test paper, he / she is entitled to one transfer with the possibility of obtaining the maximum number of points. The term of reassignment is determined by the teacher.

A student is admitted to the exam if he or she has scored at least 20 points in the semester. To receive an overall positive grade from the discipline, the exam score must be at least 24 points.

The exam is not passed if the total number of points in the discipline is less than 60 points.

Based on the PP. 4.1, 4.2 Regulations on the procedure for assessing students' knowledge, etc. 4.6.1, 7.1.5, 7.1.11, 7.1.12 Regulations on the organization of the educational process:

- the maximum number of points that can be obtained by a student during the 1st semester: 100 points;
- a student is admitted to the exam if he / she has earned at least 20 points in the semester and has received at least the minimum threshold number of points for the current and control laboratory work. To receive an overall positive grade from the discipline, the exam score must be at least 24 points.
- if during the semester the student has scored less than 60 points and is admitted to the final control, then he / she is given the opportunity to write the final control work (25/15 points), the points for which are added to the points accumulated during the semester;
- form of conducting: written work;
- types of tasks: 4 theoretical questions.

A student is admitted to the final examination in the form of an examination if he / she has earned at least 20 points for the control and laboratory work in the semester, and at the same time has scored at least 24 points. At least 60 points must be earned in order to receive an overall positive grade.

- final assessment: exam.

- the maximum number of points that can be obtained: 40 points;
- learning outcomes that will be evaluated: LO1.1, LO1.2, LO1.2, LO1.4, LO2.1, LO2.2;
- form and types of tasks: written work.

The recommended minimum is 24 points.

Типова контрольна робота складається з теоретичних та практичних завдань за матеріалом лекцій. /

Typical test work consists of theoretical and practical tasks based on lecture material.

Матеріал, що виноситься на контрольну роботу /
Material submitted for control work:

1. Класифікація інформації (об'єктів) з учителем і без учителя.
2. Опис даних і групування. Кластерний аналіз. Основні припущення. Постановка задачі.
3. Основні етапи кластерного аналізу.
4. Відстань між об'єктами і міри близькості (подібності) об'єктів.
5. Міра подібності, як інструмент кластеризації, її властивості. Побудова матриці подібності.

© Трохимчук Р.М., 2020 рік

6. Відстань між кластерами.
7. Алгоритми "найближчий сусід", "далекий сусід", компроміси між ними.
8. Критерії кластеризації (функціонали якості розбиття). Критерій мінімуму суми квадратів (мінімуму дисперсії).
9. Інші критерії, споріднені з критерієм мінімуму суми квадратів. Критерії розсіювання.
10. Матриці й критерії розсіювання.
11. Критерії кластеризації, засновані на матрицях розсіювання.
12. Алгоритми виділення кластерів: а) простий алгоритм, б) алгоритм ітеративної оптимізації.
13. Проблема вибору початкового розподілу.
14. Ієрархічне групування (дендрограма). Агломеративні (об'єднувальні) й дивізимні (подільні) методи.
15. Алгоритми сімейства FOREL.
16. Алгоритми сімейства KRAB.
17. Класифікація методів кластерного аналізу.
18. Проблема визначення кількості кластерів (класів).
19. Групування і зменшення розмірності.
20. Використання діаграми Вороного для кластеризації.
21. Методи кластеризації, що використовують теорію графів.
22. Штучна нейронна мережа Кохонена і задача кластеризації, машина Больцмана.
23. Аналіз та інтерпретація результатів кластерного аналізу.
24. Методи візуалізації результатів кластерного аналізу.

/

1. Classification of information (objects) with and without the teacher.
2. Data description and grouping. Cluster analysis. Basic assumptions. Formulation of the problem.
3. The main stages of cluster analysis.
4. The distance between objects and the degree of similarity (similarity) of objects.
5. Similarity measure as a clustering tool, its properties. Construction of a similarity matrix.
6. Distance between clusters.
7. Algorithms "nearest neighbor", "distant neighbor", compromises between them.
8. Clustering criteria (partition quality functionals). The criterion for the minimum of the sum of squares (the minimum of variance).
9. Other criteria related to the minimum sum of squares criterion. Scattering criteria.
10. Matrices and scattering criteria.
11. Clustering criteria based on scattering matrices.
12. Algorithms for cluster allocation: a) simple algorithm, b) iterative optimization algorithm.
13. The problem of choosing the initial distribution.
14. Hierarchical grouping (dendrogram). Agglomerative (unifying) and divisive (divisive) methods.
15. FOREL family algorithms.
16. KRAB family algorithms.
17. Classification of cluster analysis methods.
18. The problem of determining the number of clusters (classes).
19. Grouping and dimension reduction.
20. Using the Voronoi diagram for clustering.
21. Clustering methods using graph theory.
22. Kohonen artificial neural network and clustering problem, Boltzmann machine.
23. Analysis and interpretation of cluster analysis results.
24. Methods of visualizing the results of cluster analysis.

7.2 Організація оцінювання:

Терміни проведення форм оцінювання:

1. Контрольна робота: до 14 тижня семестру.
2. Захист самостійної роботи: до 14 тижня семестру.

Студент має право на одне перескладання кожної контрольної роботи із можливістю отримання максимально 80% початково визначених за цю контрольну роботу балів. Термін перескладання визначається викладачем.

У випадку відсутності студента з поважних причин відпрацювання та перездачі контрольних робіт здійснюються у відповідності до „Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу” від 1 жовтня 2010 року.

/

Terms of evaluation forms:

1. Test Work: up to 14 weeks of the semester.
2. Defense of independent work: up to 14 weeks of the semester.

The student has the right to one retake of each test with the possibility of obtaining a maximum of 80% of the points initially determined for this test. The term of reassembly is determined by the teacher.

In case of absence of a student for valid reasons working off and transfer of tests are carried out according to "Regulations on the order of an estimation of knowledge of students at the credit-modular system of the organization of educational process" from October 1, 2010.

7.3. Шкала відповідності оцінок / Rating scale

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

8. Структура навчальної дисципліни. Тематичний план занять / Structure of the discipline. Thematic lesson plan

№ Лекції/ No lectures	Назва лекції (теми) / Title of the lecture (topic)	Кількість годин / Number of hours	
		Лекції/ Lectures	Сам. робота/ Ind. work
1	Тема 1. Класифікація інформації (об'єктів) з учителем і без учителя. Опис даних і групування. Кластерний аналіз. Основні припущення. Постановка задачі. Classification of information (objects) with and without the teacher. Data description and grouping. Cluster analysis. Basic assumptions. Formulation of the problem.	2	6
2	Тема 2. Основні етапи кластерного аналізу. Відстань між об'єктами і міри близькості (подібності) об'єктів. The main stages of cluster analysis. The distance between objects and the degree of similarity (similarity) of objects.	3	8
3	Тема 3. Міра подібності, як інструмент кластеризації, її властивості. Побудова матриці подібності. Відстань між кластерами. A measure of similarity as a clustering tool, its properties. Construction of a similarity matrix. Cluster distance.	2	6
4	Тема 4. Алгоритми "найближчий сусід", "далий сусід",	2	6

	компроміси між ними. / Algorithms "nearest neighbor", "far neighbor", compromises between them.		
5	Тема 5. Критерії кластеризації (функціонали якості розбиття). Критерій мінімуму суми квадратів (мінімуму дисперсії). Інші критерії, споріднені з критерієм мінімуму суми квадратів. / Clustering criteria (partition quality functionals). The criterion for the minimum of the sum of squares (the minimum of variance). Other criteria related to the minimum sum of squares criterion.	3	6
6	Тема 6. Критерії розсіювання. Матриці й критерії розсіювання. / Scattering criteria. Matrices and scattering criteria.	2	6
7	Тема 7. Критерії кластеризації, засновані на матрицях розсіювання. / Clustering criteria based on scattering matrices.	2	6
8	Тема 8. Алгоритми виділення кластерів: а) простий алгоритм, б) алгоритм ітеративної оптимізації. Проблема вибору початкового розподілу. / Clustering algorithms: a) simple algorithm, b) iterative optimization algorithm. The problem of choosing the initial distribution.	3	6
9	Тема 9. Ієрархічне групування (дендрограма). Агломеративні (об'єднувальні) й дивізійні (подільні) методи. / Hierarchical grouping (dendrogram). Agglomerative and divisive methods.	3	6
10	Тема 10. Алгоритми сімейства FOREL. / Family of FOREL algorithms.	2	6
11	Тема 11. Алгоритми сімейства KRAB. / Family of KRAB algorithms.	2	6
12	Тема 12. Класифікація методів кластерного аналізу. Проблема визначення кількості кластерів (класів). / Classification of cluster analysis methods. The problem of determining the number of clusters (classes).	2	6
13	Тема 13. Групування і зменшення розмірності. Використання діаграми Вороного для кластеризації. / Grouping and Dimensioning. Using the Voronoi diagram for clustering.	3	6
14	Тема 14. Методи кластеризації, що використовують теорію графів. / Clustering methods using Graph Theory.	2	6
15	Тема 15. Штучна нейронна мережа Кохонена і задача кластеризації, машина Больцмана. / Artificial neural network Kohonen and clustering problem, Boltzmann machine.	3	6
16	Тема 16. Аналіз та інтерпретація результатів кластерного аналізу. / Analysis and interpretation of cluster analysis results.	2	4
17	Тема 17. Методи візуалізації результатів кластерного аналізу. / Methods for visualizing the results of cluster analysis.	2	6
Контрольна робота / Control work		2	4
ВСЬОГО / TOTAL		42	106

Лекцій – 42 год., / Lectures – 42 hours,

Самостійна робота – 106 год., / Independent work – 106 hours,

Консультації – 2 год. / Consultation – 2 hours

9. Рекомендовані джерела: / Recommended sources:

Основні / Basic:

1. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.
2. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
3. Leskovec J. Mining of Massive Datasets / Jure Leskovec Anand Rajaraman, Jeffrey David Ullman // Stanford Univ. — 2010.
4. Data is the New Oil By Michael Palmer http://ana.blogs.com/maestros/2006/11/data_is_the_new.html

Додаткові / Additional:

1. Berry, Michael J. A. “Data mining techniques: for marketing, sales, and customer relationship management” / Michael J.A. Berry, Gordon Linoff. — 2nd ed.
2. Larose, Daniel T. “Discovering knowledge in data: an introduction to data mining” / Daniel T. Larose
3. J. Ross Quinlan. C4.5: Programs for Machine learning. Morgan Kaufmann Publishers 1993.
4. S.Murthy. Automatic construction of decision trees from data: A Multi- disciplinary survey. 1997.
5. W. Buntine. A theory of classification rules. 1992.
6. Machine Learning, Neural and Statistical Classification. Editors D. Mitchie et.al. 1994.
7. Ganti V., Gerhke J., Ramakrishan R. CACTUS – Clustering Categorical Data Using Summaries. In Proc KDD'99, 1999