

Моделі, інформаційні технології та архітектура систем обробки демографічної інформації



ЧЕРТОВ

Олег

Романович



к.т.н., доцент кафедри прикл. математики
НТУУ "Київський політехнічний інститут"



консультант Світового банку, проект P076338



IEEE Member



IACSIT Member

Науковий консультант -
МОЛЧАНОВ
Олександр
Артемійович,
д.т.н., проф.



Демографічна інформація

- ▼ **Демографічна інформація** — дані про чисельність населення, його структуру, розміщення та динаміку [Пальян, підручник].
- ▼ Структура населення досліджується:
 - за **демографічними** ознаками — стать, вік, шлюбний стан тощо;
 - за **соціальними** ознаками — громадянство, етнічне походження, рідна мова, джерела засобів існування, освіта тощо.



Потреба в показниках статистики населення

- ▼ від **владних структур** — для розрахунку податків з державного бюджету на соціальні потреби населення і визначення міжбюджетних трансфертів в галузях охорони здоров'я, соціального захисту населення, культури, міського громадського транспорту тощо
- ▼ від **наукових установ** — для побудови довгострокових і короткострокових демографічних, соціологічних та економічних прогнозів
- ▼ від **комерційних підприємств** — при формуванні планів свого подальшого розвитку та проведенні маркетингової діяльності
- ▼ від **широких верств населення** країни



Основне джерело демографічної інформації

- ▼ Основним джерелом демографічної інформації є **перепис населення**, а в міжпереписний період — вибіркові дослідження та результати оцінок населення
- ▼ Характеристики перепису як статистичного спостереження:
 - індивідуальний підрахунок респондентів
 - одночасність
 - усталена періодичність
 - суцільність



Специфіка систем обробки демографічної інформації

країна	Україна (2001 р.)	Росія (2002 р.)
особливість		
територіально-ієрархічний розподіл системи	27 регіон. центрів	67 регіон. центрів
підтримка дуже великих обсягів даних	≈ 68 млн. переписних форм	≈ 220 млн. переписних форм
підтримка великої кількості різномірних правил контролю первинних та зведених даних	≈ 200 тисяч правил	≈ 200 тисяч правил

- ▼ **Перепис** — це найзатратніша операція, котру країна проводить в мирний час [Monty Wood, прес-секретар Бюро перепису населення США]



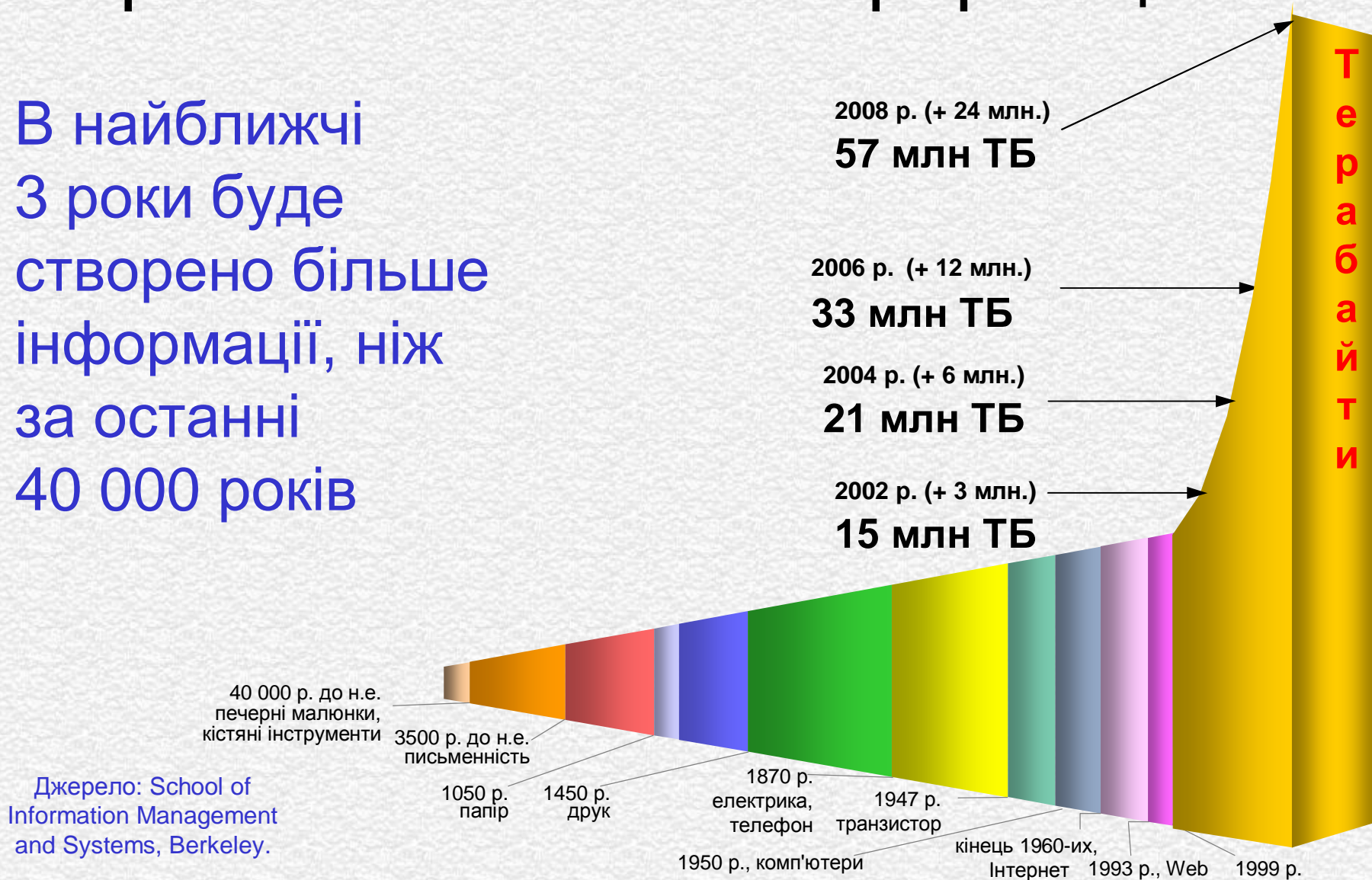
Науково-прикладна проблема, розв'язувана в дисертаційній роботі

- ▼ Мають бути розроблені **моделі, інформаційні технології** та **архітектурні рішення**, які
 - 1) дозволять розширити функціональні можливості сучасних систем з обробки демографічної інформації,
 - 2) гарантуватимуть знеособленість та анонімність даних про респондентів.



Зростання обсягів інформації

В найближчі
3 роки буде
створено більше
інформації, ніж
за останні
40 000 років



Джерело: School of
Information Management
and Systems, Berkeley.



Вік	Всього	у тому числі	
		чоловіки	жінки
Все населення	48240902	22316317	25924585
у віці:			
0-4 р.	1974195	1012224	961971
5-9 р.	2559107	1311943	1247164
10-14 р.	3416561	1750539	1666022
15-19 р.	3891568	1989538	1902030
20-24 р.	3489588	1766985	1722603
25-29 р.	3402010	1700516	1701494
30-34 р.	3204103	1586043	1618060
35-39 р.	3417079	1660397	1756682
40-44 р.	3828331	1833438	1994893
45-49 р.	3470419	1623194	1847225
50-54 р.	3182588	1454550	1728038
55-59 р.	2062711	892852	1169859
60-64 р.	3364050	1398332	1965718
65-69 р.	2158171	868633	1289538
70-74 р.	2239064	806587	1432477
75-79 р.	1500920	416155	1084765
80-84 р.	618819	147952	470867
85-89 р.	324697	65475	259222
90-94 р.	100593	18643	81950
95-99 р. 2011-2010	15119	2419	12700
100 років і			

Регламентні таблиці інформаційно-облікових систем

Україна, перепис 2001 р.:

- 6852 розрізи вихідних таблиць
- 17 друкованих збірок:



	м. Київ	Донецька	Сумська
українська	10 000	2000	1 000
російська	5000	500	250
білоруська	5000	500	250

Запити в інформаційно-аналітичних системах

- ▼ **Верховна Рада України:** соціально-демографічна характеристика циган
- ▼ **ЮНЕСКО:** населення не молодше 70 років за рівнем освіти
- ▼ **ООН:** населення за громадянством та віком
- ▼ **Статкомітет СНД:** мігранти, котрі володіють російською мовою, за місцем народження
- ▼ **Роскомстат:** населення, котре, володіє російською мовою, за віком
- ▼ Аналіз наявних даних шляхом побудови розрізу багатовимірного кубу
- ▼ Але як віднайти неявні закономірності розподілу даних?
- ▼ Наприклад, як відібрати основні демографічні та соціально-економічні чинники, що сприятимуть підвищенню народжуваності в країні?
- ▼ Потрібні засоби інтелектуального аналізу даних


Визначення Data mining

- ▼ **Інтелектуальний аналіз даних** (Data mining) — нетривіальний процес виявлення коректних, раніш невідомих, потенційно корисних і придатних для інтерпретації закономірностей в даних [Piatetsky-Shapiro, 1992]
- ▼ Сфера застосування:
 - **наукові дослідження** — пошук маркерів в молекулярній генетиці та розшифровка генома людини, аналіз складних хімічних сполук
 - **інженерна практика** — діагностичний контроль технічного стану обладнання
 - **бізнес** — скоринг, аналіз кошика покупця в роздрібній торгівлі, виявлення лояльності користувачів мобільним зв'язком
- ▼ Методи інтелектуального аналізу даних в промислових системах для обробки демографічної інформації наразі не застосовуються взагалі



IPUMS-International

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA



Home | Variables | Create Extract | FAQ | Contact Us | Login

PROJECT	<h2>Integrated Public Use Microdata Series International</h2> <p>census microdata for social and economic research</p> <hr/> <p>IPUMS-International is a project dedicated to collecting and distributing census data from around the world. Its goals are to:</p> <ul style="list-style-type: none"> • Collect and preserve data and documentation • Harmonize data • Disseminate the data absolutely free! <hr/> <p>55 countries - 159 censuses - 326 million person records</p>	IPUMSI News
<p>About IPUMS-I</p> <p>How to Cite IPUMS-I</p> <p>User Registration and Login</p>		<p>June 2010 data release</p> <p>2009 award winners</p> <p>Mortality and fertility data</p> <p>Improved extract system</p> <p>NIH extends IPUMS-I</p> <p>All news items</p>
DATA		MPC Data Projects
<p>Variables</p> <p>Create Data Extract</p> <p>Download Data Extract</p> <p>GIS and Other Data Files</p>		<p>IPUMS-USA and others</p>
SAMPLES		
<p>Sample Descriptions</p> <p>Variance Estimation</p> <p>Source Documents</p>		

55

країн,

159

переписів,

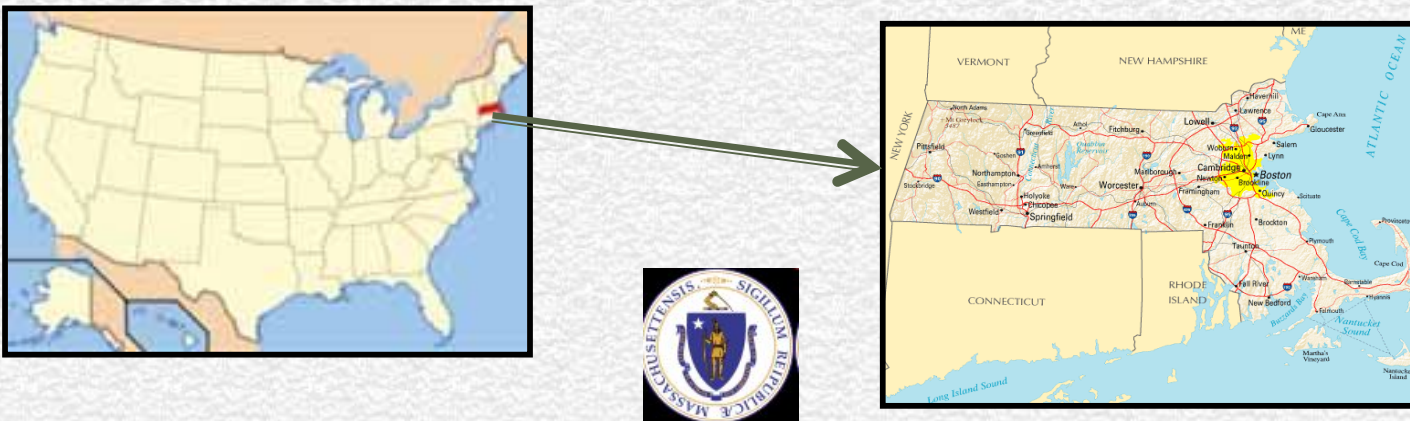
326 млн.

персон.
записів

<https://international.ipums.org/international/>



Забезпечення індивід. анонімності



Массачусетс, населення: 6,35 млн. (2000 р.)

Sweeney (2001), % виборців штату:

- 12% з унікальною датою народження (день, місяць, рік),
- 29% з унік. датою народження та статтю,
- 69% з унік. датою народження та 5-ти цифр. поштовим індексом,
- **87%** з унік. датою народження, статтю та індексом,
- **97%** з унік. датою народження та 9-ти цифр. поштовим індексом.



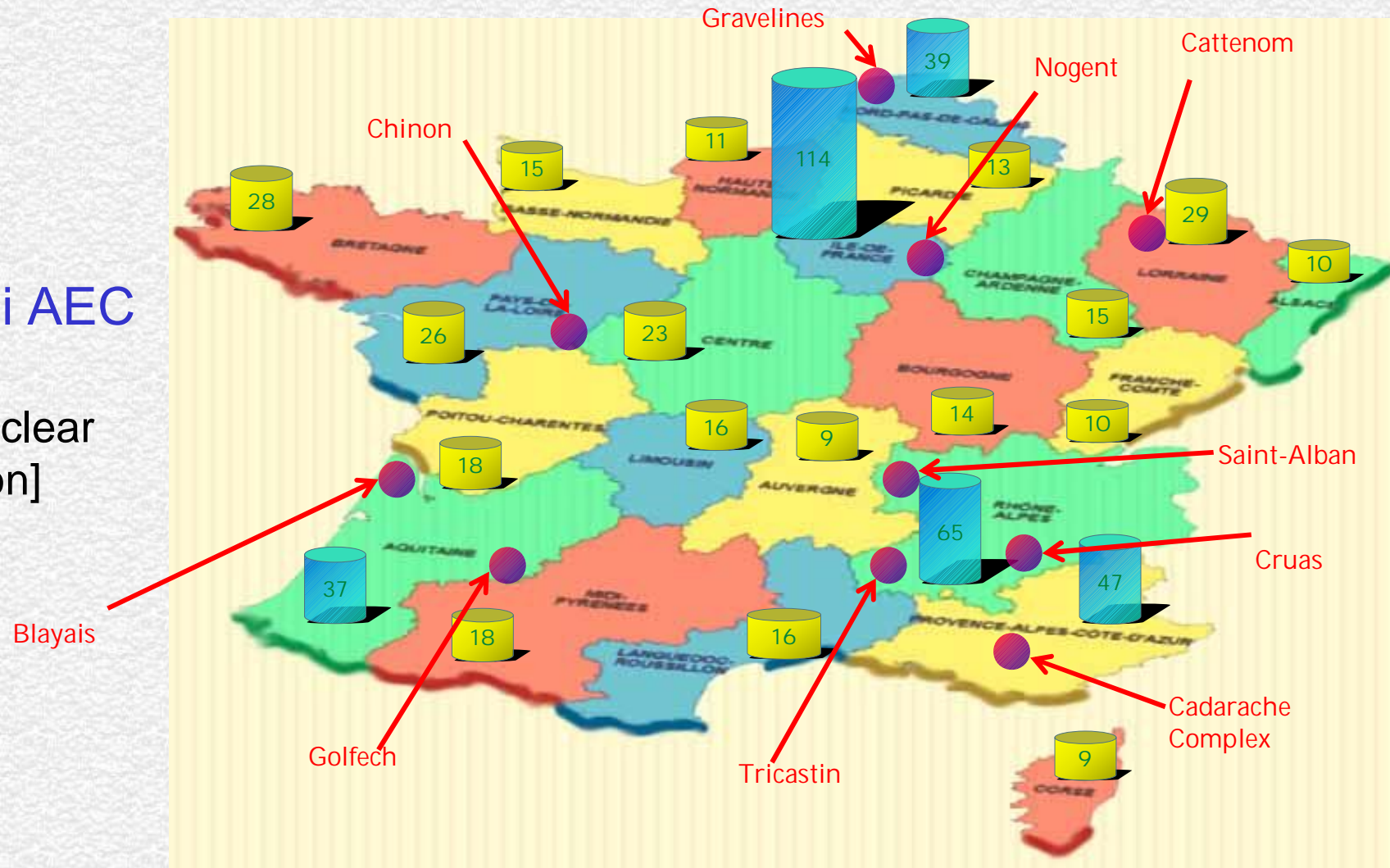
Способи приховування значень атрибутів

- ▼ Ідентифікатори ==> **шифрування чи вилучення.**
- ▼ Для інших видів атрибутів ==> **маскування:**
 - **пертурбативні (зі збуренням)** - початкова множина даних викривляється:
 - мікроагрегація (заміна кожного запису на центроїд його групи)
 - обмін даними (перестановка значень конфіденційних атрибутів поміж записами)
 - додавання шуму
 - ...
 - **непертурбативні:**
 - приховування (вилучення деяких даних із початкової множини)
 - узагальнення (укрупнення даних)
 - ...
- ▼ **k-анонімність**, тобто кожній комбінації значень певних атрибутів повинно відповідати, принаймні, k записів, що мають таку ж комбінацію.



Забезпечення групової анонімності

Виділені АЕС Франції
[World Nuclear Association]



Об'єкт і предмет дослідження

- ▼ **Об'єкт дослідження:** процеси накопичування, обробки та збереження демографічної інформації, що породжують проблему необхідності розширення функціональних можливостей інформаційних систем, які реалізують зазначені процеси
- ▼ **Предмет дослідження:** моделі, інформаційні технології та архітектурні рішення, які забезпечують організацію обробки демографічної інформації



Корпоративна архітектура

- ▼ “**Корпоративна архітектура** пояснює спосіб, за допомогою якого загальне бачення діяльності організації відображається в його структурі та динаміці. На різних рівнях абстракції вона дає єдиний набір моделей, принципів, керівництв і політик, котрі використовуються для трансформації, синхронізації та розвитку систем у масштабі і контексті діяльності всієї організації в цілому” [Global Enterprise Architecture Organization]
- ▼ Відомча **функціонально-інформаційна архітектура** = корпоративна архітектура конкретного відомства чи галузі



Основні рівні корпоративної архітектури

№	Назва рівня	Пояснення
1	Архітектура діяльності	визначає бізнес-процеси та організаційні структури організації
2	Архітектура інформації	описує логічну та фізичну організацію даних, ресурси з управління даними, санкціонування доступу до них
3	Архітектура застосувань	надає інтерфейси взаємодії прикладних систем між собою, з зовнішніми системами, джерелами чи користувачами даних
4	Технологічна архітектура (інфраструктура)	включає апаратні засоби, системне програмне забезпечення, мережі, комунікаційні протоколи і т.д.

Згідно [Togaf Ver. 9 Enterprise Edition, 2009;
FEA Consolidated Reference Model, Ver. 2.3, 2007]



Адаптована модель Захмана (Framework for Information Systems Architecture)

Домени Рівні абстракції	ЧОМУ (мотивація)	ЩО (дані)	ЯК (функції)	ДЕ (дислокація)	ХТО (люди)	КОЛИ (час)
Контекст, рівень керівництва	законодавчо визначені цілі	статистичні обстеження та показники	перелік послуг і процесів	3-х рівневе розташування органів статистики, місця проведення досліджень	споживачі інформації	програма статистичних досліджень
Концептуальна модель, рівень методологів	моделі показників звітності	метамоделі, класифікатори	модель статистичних процесів	схема логістики	моделі потоків робіт, перелік ролей	календар і періодичність спостереження
Логічна модель, рівень архітекторів	суб'єкти та моделі діяльності	логічні моделі даних	архітектура застосувань	модель розподіленої архітектури	модель ролевих відносин	діаграма подій, їхня синхронізація
Технолог. модель, рівень проєктувальників	посадові інструкції	фізичні моделі даних	специфікації використання	інфраструктура	інтерфейси користувача	специфікація подій
Деталі реалізації, рівень виконавців	керівництва користувача, регламент робіт	опис структур даних	програмний код	мережна архітектура	автоматизовані робочі місця	визначення часових прив'язувань
Практика застосування	показники точності, надійності інформації	первинні та агреговані дані, мікродані	регламенти, що реалізуються, послуги, що надаються	фізичне розташування обладнання	постачальники та споживачі інформації і послуг	розклад надання (чи доступності) інформації



Місце рівня “системної архітектури”

1	Архітектура діяльності	визначає бізнес-процеси та організаційні структури організації
	Системна архітектура	описує роботу системи з даними
2	Архітектура інформації	описує логічну та фізичну організацію даних, ресурси з управління даними, санкціонування доступу до них

Концептуальна модель, рівень методологів	моделі показників звітності	метамоделі, класифікатори	модель статистичних процесів	схема логістики	моделі потоків робіт, перелік ролей	календар і періодичність спостереження
Системна модель, рівень методологів і архітекторів	пошук інформації	метамоделі, логічні моделі даних	робота з даними, доступ до них	модель системної архітектури	моделі запитів	узгодженість і синхронізація запитів
Логічна модель, рівень архітекторів	суб'єкти та моделі діяльності	логічні моделі даних	архітектура застосувань	модель розподіленої архітектури	модель рольових відносин	діаграма подій, їхня синхронізація



Види (рівні) системної архітектури

риса	OLTP	OLAP	Data mining
характер (рівень) даних	головним чином, первинні	в основному, консолідовані	первинні та консолідовані
мінливість даних	висока (з кожною транзакцією)	низька	низька
типова операція	зберігання та зміна інформації	аналіз наявних даних	пошук неявних закономірностей
звіти	регламентні	нерегламентні (з визначеним переліком атрибутів)	нерегламентні (з динамічним переліком атрибутів)
дані, що оброблюються	поточні	історичні та поточні	історичні та поточні
базова структура	таблиця / первинний ключ	куб / вимір	кластер, клас, правило
пріоритет	продуктивність	гнучкість	інтелектуальність



Напрями застосування інтелектуального аналізу даних в статистиці

1. Класичні методи інтелектуального аналізу даних
2. Методи пошуку заздалегідь визначеного шаблону в наявних даних



Підвищення народжуваності в Каліфорнії

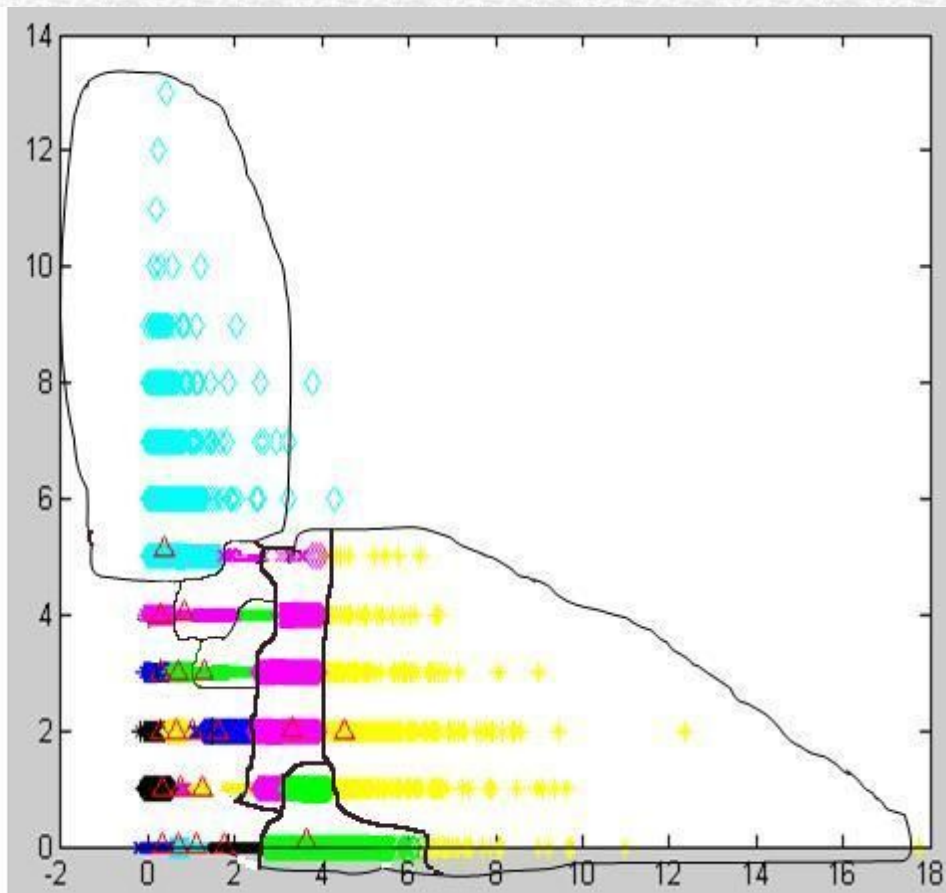
- ▼ Від'ємне сальдо внутрішньої міграції = число виїхавших перевищує число прибулих більше, ніж на 1,3 млн. осіб (з 2000 р. по 2008 р.).
- ▼ Чинники, що впливають на підвищення народжуваності (?):
 - матеріальні статки родини
 - кількість уже наявних дітей в сім'ї
 - етнічне походження подружжя
- ▼ Каліфорнія, 5-відсоткові мікродані перепису населення США, 2000 р.
- ▼ Вибірка (112347 осіб): сім'ї з працездатним подружжям, обоє віком від 18 до 50 років.



Класифікація множини даних (20 класів)

кількість
дітей в сім'ї

5: 15
 4: 8, 20
 3: 7, 12, 6
 2: 10, 18, 13, 1, 14, 11
 1: 16, 2, 5
 0: 19, 9, 3, 4, 17



нормований сукупний дохід
родини за 1999 р.

- ▼ трикутники позначають центри класів
- ▼ до двокрапки вказана середня кількість дітей в родинах відповідних класів
- ▼ класи 6 і 20 – шукані!

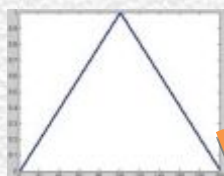
Напрями застосування інтелектуального аналізу даних в статистиці

1. Класичні методи інтелектуального аналізу даних
2. Методи пошуку заздалегідь визначеного шаблону в наявних даних (напрям **template matching** в розпізнаванні образів):

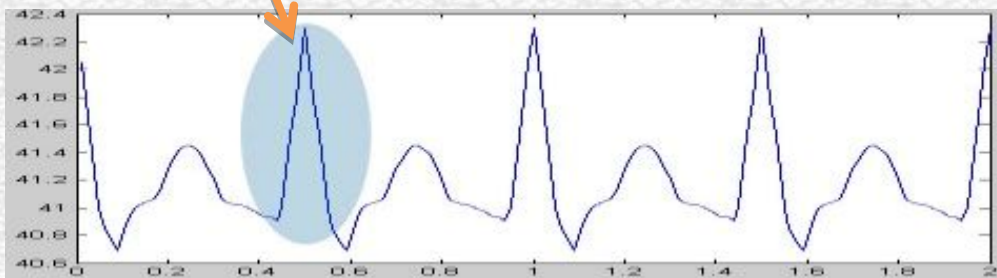


Пошук шаблону (template matching)

- φ – шаблонна функція (шаблон), екземпляр якої ми хочемо віднайти в сигналі



Шаблон φ



Сигнал s із екземпляром шаблону

- Апроксимація сигналу лінійною комбінацією функцій певного базису:
 - зокрема, поліномами Кунченка:
 - базис як лінійно-незалежні перетворення шаблону:

$$f_1(\varphi), f_2(\varphi), \dots, f_n(\varphi), \dots$$
 - поліном наближення:

$$P_n = \sum_{i=1}^n c_i f_i(\varphi)$$

- s – сигнал зі скінченною енергією, який ми хочемо проаналізувати

Поліноми Ю.П. Кунченка

Коефіцієнти c_k знаходяться як розв'язок системи лінійних рівнянь:

$$\sum_{k=1}^n c_k F_{i,k} = F_{i,b}, \quad i = \overline{1, n},$$

де $F_{i,j} = \langle f_i(\varphi), f_j(\varphi) \rangle - \frac{\langle f_i(\varphi), f_0(\varphi) \rangle \cdot \langle f_j(\varphi), f_0(\varphi) \rangle}{\langle f_0(\varphi), f_0(\varphi) \rangle},$

$$\langle f_i(\varphi), f_j(\varphi) \rangle = \int_a^b f_i(\varphi) \cdot f_j(\varphi) dx$$

При цьому $f_b(\varphi) \equiv s.$

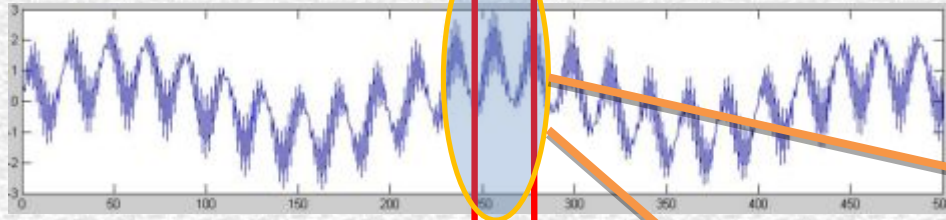
Коефіцієнт ефективності:
$$e = \frac{\sum_{k=1}^n c_k F_{k,b}}{\langle f_b(\varphi), f_b(\varphi) \rangle}$$



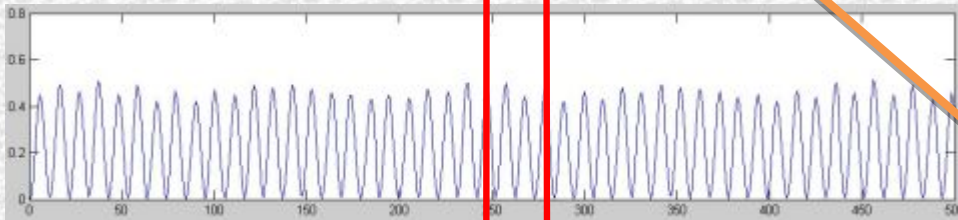


$$\varphi = 1 - |x|, [-1; 1]$$

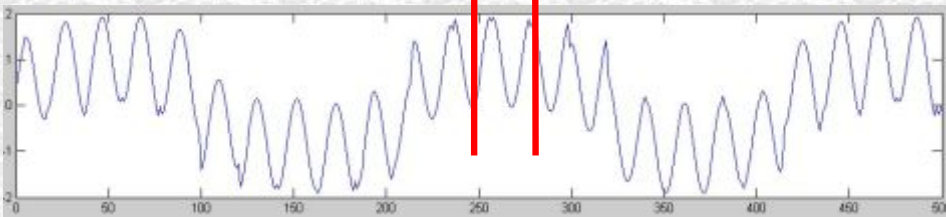
Модельний приклад пошуку шаблону в сигналі



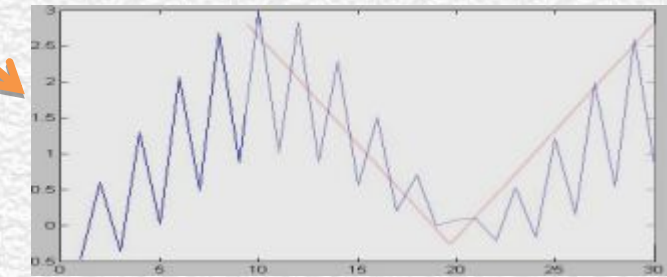
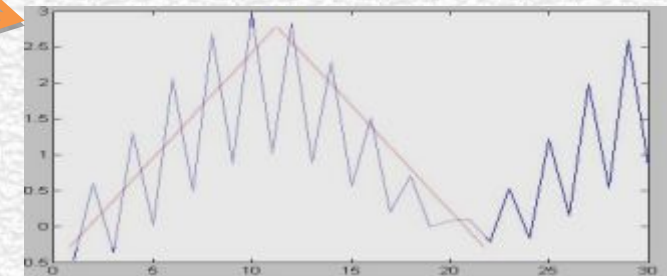
s із накладеним шаблоном



ефектограма



поліном Кунченка (наближення)

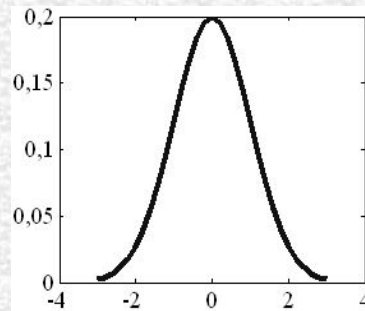
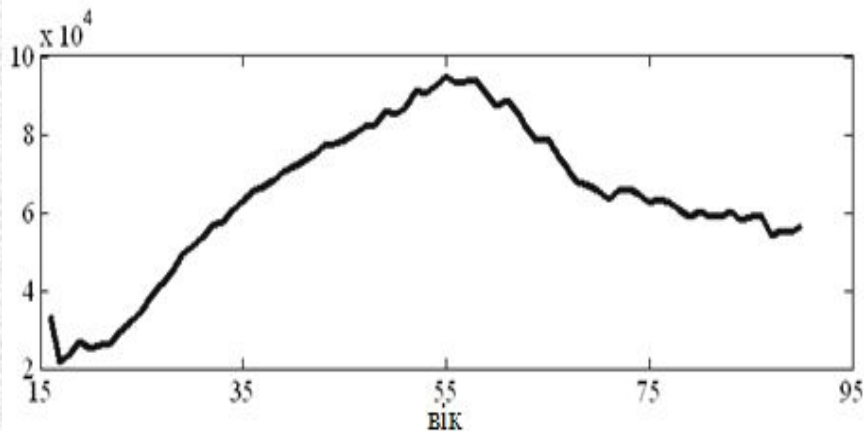


Піки в ефектограмі – відліки сигналу, з яких починається шаблон. Другий пік у вікні – інвертований шаблон.



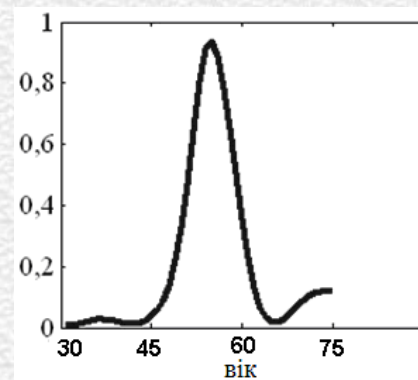
Приклад на реальних даних (Каліфорнія, 2000 р.)

Розподіл середнього доходу домогосподарства в залежності від віку голови

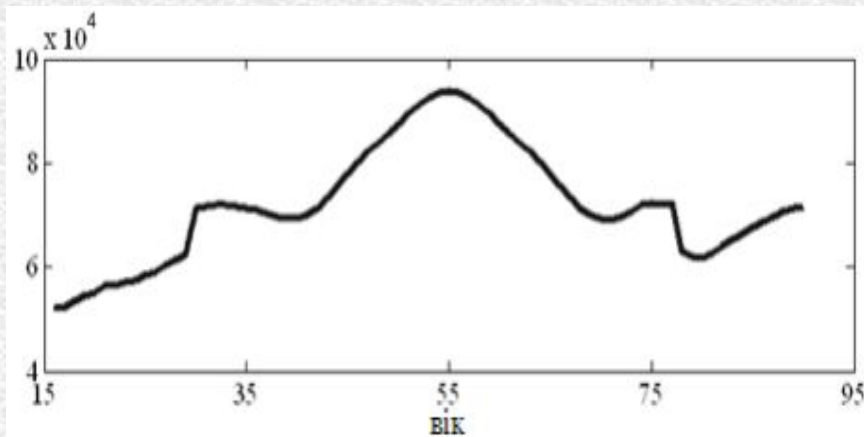


шаблон
(гаусіана):

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



коефіцієнти
ефективності,
макс. = 0,94
при 55 роках



Поліном наближення Кунченка:
 $107 \cdot (0,0068 + 0,0272\psi(x) + 0,2071\psi^2(x) + 0,8779\psi^3(x) - 1,7866\psi^4(x) + 1,4059\psi^5(x))$

Адаптована модель Захмана для статистичної галузі України

Домени Рівні абстракції	ЧОМУ (мотивація)	ЩО (дані)	ЯК (функції)	ДЕ (дислокація)	ХТО (люди)	КОЛИ (час)
Контекст, рівень керівництва	законодавчо визначені цілі	статистичні обстеження та показники	перелік послуг і процесів	3-х рівневе розташування органів статистики, місця проведення досліджень	споживачі інформації	програма статистичних досліджень
Концептуальна модель, рівень методологів	моделі показників звітності	метамоделі, класифікатори	модель статистичних процесів	схема логістики	моделі потоків робіт, перелік ролей	календар і періодичність спостереження
Логічна модель, рівень архітекторів	суб'єкти та моделі діяльності	логічні моделі даних	архітектура застосувань	модель розподіленої архітектури	модель рольових відносин	діаграма подій, їхня синхронізація
Технолог. модель, рівень проєктувальників	посадові інструкції	фізичні моделі даних	специфікації використання	інфраструктура	інтерфейси користувача	специфікація подій
Деталі реалізації, рівень виконавців	керівництва користувача, регламент робіт	опис структур даних	програмний код	мережна архітектура	автоматизовані робочі місця	визначення часових прив'язувань
Практика застосування	показники точності, надійності інформації	первинні та агреговані дані, мікродані	регламенти, що реалізуються, послуги, що надаються	фізичне розташування обладнання	постачальники та споживачі інформації і послуг	розклад надання (чи доступності) інформації



Загальна модель стат. бізнес-процесів (2009 р.)

1 Уточнення потреб	2 Проектування	3 Формування	4 Збір	5 Обробка	6 Аналіз	7 Поширення	8 Архів	9 Оцінка
1.1 Визначення потреб у інформації	2.1 Проектування результатів	3.1 Створення інструменту збору даних	4.1 Вибірка	5.1 Інтеграція даних	6.1 Підготовка проекту вихідних даних	7.1 Оновлення вихідних систем	8.1 Визначення правил архівування	9.1 Збір вхідних даних для оцінки
1.2 Консультації та підтвердження потреб	2.2 Проектування описів змінних	3.2 Створення (збільшення) компонентів процесу	4.2 Початок збору	5.2 Класифікація і кодування	6.2 Підтвердження вихідних даних	7.2 Виробництво продуктів поширення	8.2 Ведення архівного сховища	9.2 Проведення оцінки
1.3 Визначення завдань вихідних даних	2.3 Проектування методів збору даних	3.3 Формування робочих процесів	4.3 Збір даних	5.3 Перегляд, підтвердження і редагування	6.3 Дослідження та пояснення	7.3 Випуск продуктів поширення	8.3 Збереження даних і метаданих	9.3 Узгодження плану дій
1.4 Визначення понять	2.4 Проектування методів визначення меж та вибірки	3.4 Випробування системи виробництва	4.4 Завершення збору	5.4 Використання розрахункових даних	6.4 Контроль над відкриттям інформації	7.4 Просування продуктів поширення	8.4 Знищення даних і відповідних метаданих	
1.5 Перевірка доступності даних	2.5 Проектування методів обробки даних	3.5 Випробування статистичного бізнес-процесу		5.5 Отримання нових змінних та статистичних одиниць	6.5 Оформлення вихідних даних	7.5 Здійснення підтримки користувачів		
1.6 Підготовка бізнес-справи	2.6 Проектування робочого процесу	3.6 Завершення формування системи виробництва		5.6 Обчислення значень				



Загальна модель та територіальні рівні статистичної обробки в Україні



Структура мікрофайлу

№ д/г	перс. №	стать	вік	сім. стан	місце прож.	професія
3413	1	2	45	3	100	40
3413	2	1	13	2	100	0
3413	3	2	9	2	100	0
3485	1	1	25	1	100	117
3485	2	2	22	1	100	117
3498	1	2	76	4	200	0
3501	1	1	48	1	200	254
3501	2	2	47	1	200	962
3501	3	1	18	2	200	120

■ Сутнісні атрибути

■ Параметризує атрибут – за яким йде розподіл даних

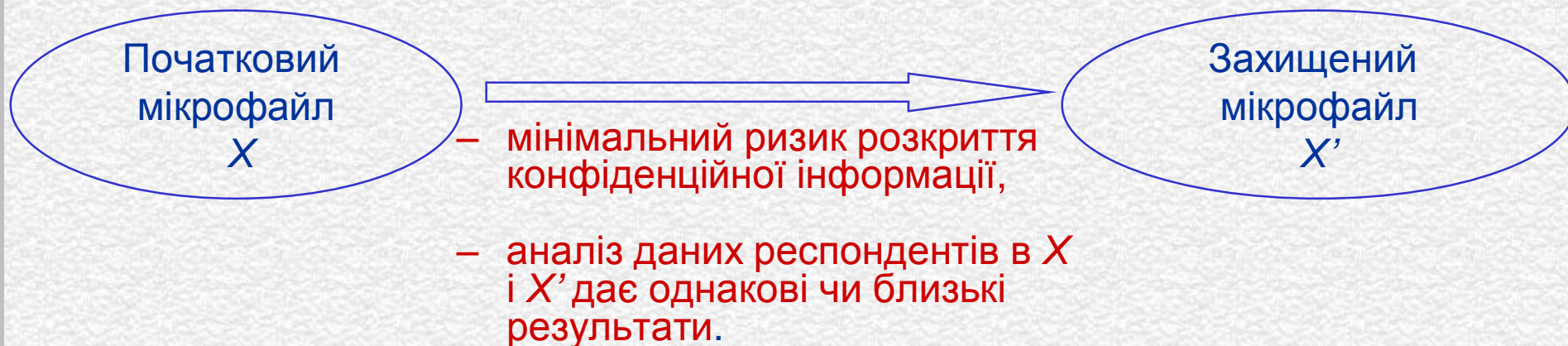
■ Комбінація значень сутнісних атрибутів:

(1, 25) і (1, 18)

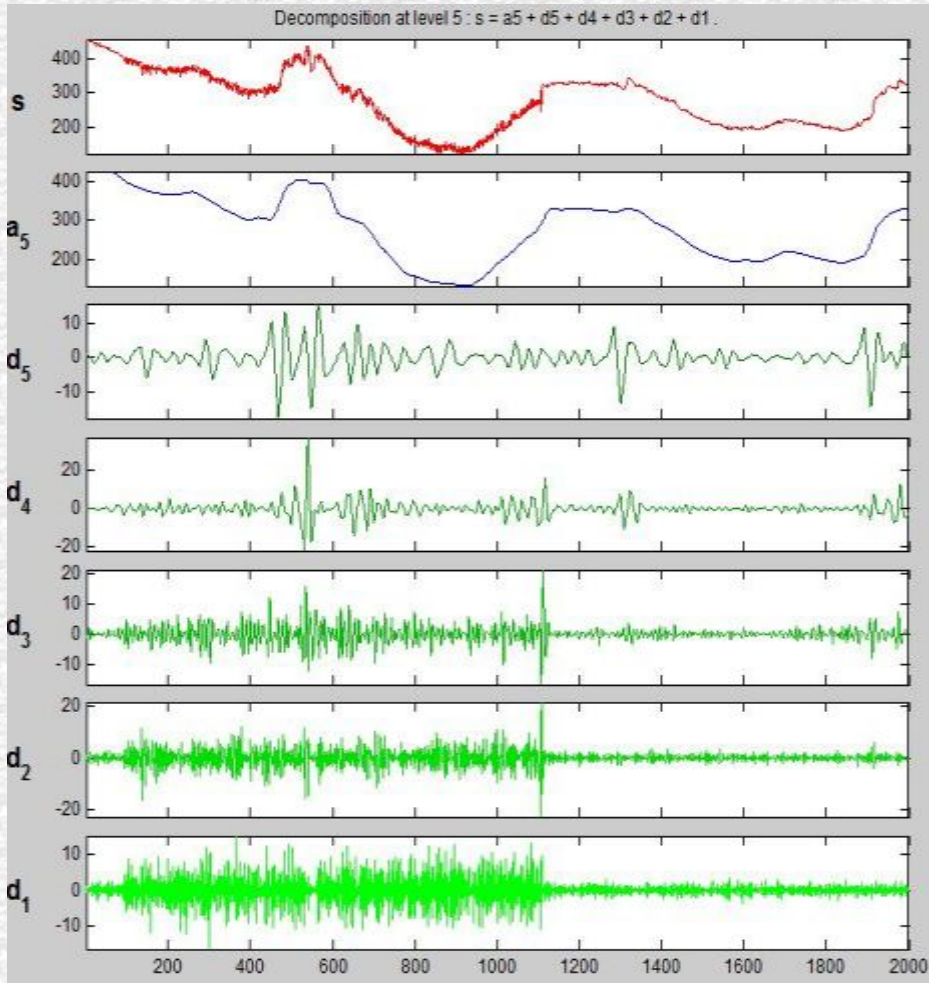
Задача забезпечення групової анонімності даних

Мікрофайл – набір даних, зведених в один файл, що містить атрибуtnі записи на кожного з респондентів.

Забезпечення **групової анонімності даних** – це процес такого перетворення початкового набору даних, в результаті чого унеможлиблюється з'ясування правильного розподілу по значенням параметризуючого атрибуту сутнісних комбінацій значень певної групи респондентів.



Приклад дискретного діадного вейвлет-розкладення сигналу



▼ Початковий сигнал

▼ Низькочастотний рівень 5
(1/16 початкових відліків)

апроксимація

▼ Високочастотний рівень 5
(1/16 початкових відліків)

▼ Високочастотний рівень 4
(1/8 початкових відліків)

▼ Високочастотний рівень 3
(1/4 початкових відліків)

▼ Високочастотний рівень 2
(1/2 початкових відліків)

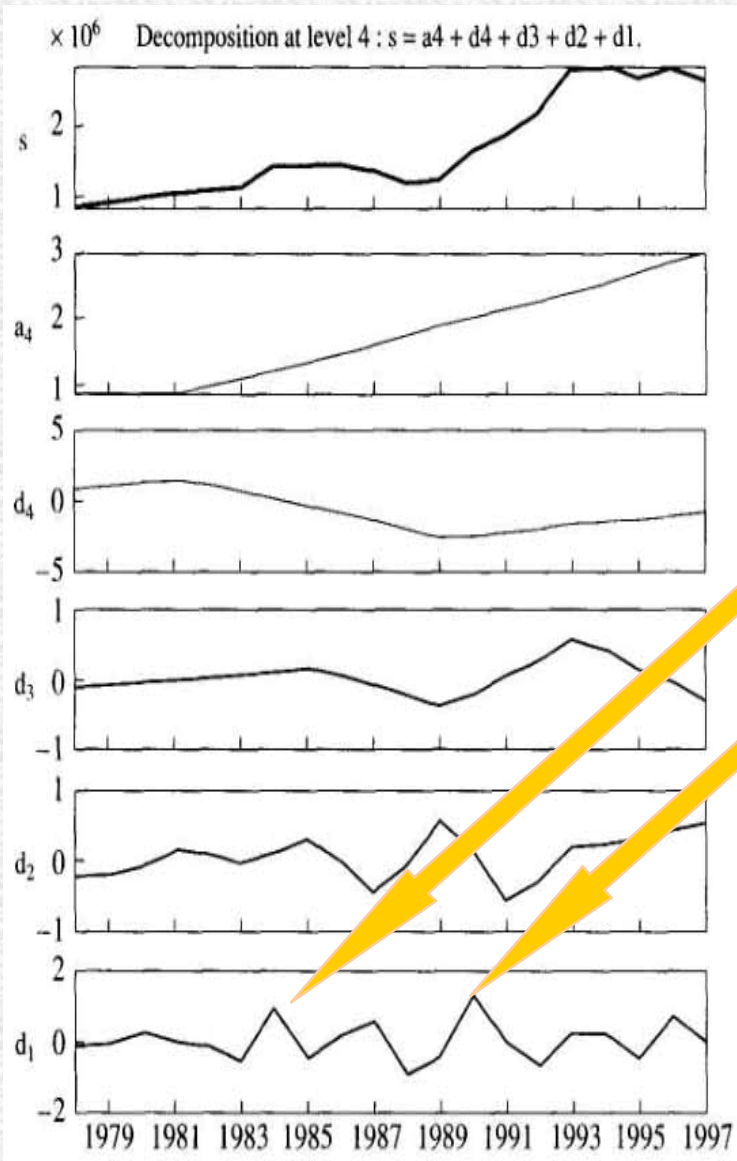
▼ Високочастотний рівень 1
(всі початкові відліки)

деталізація

Зареєстровані кримінальні злочини в СРСР

▼“Деталізуючі складові відображають приховані особливості часових рядів, важливі для коротко- та середньо-термінових прогнозів соціальних процесів”

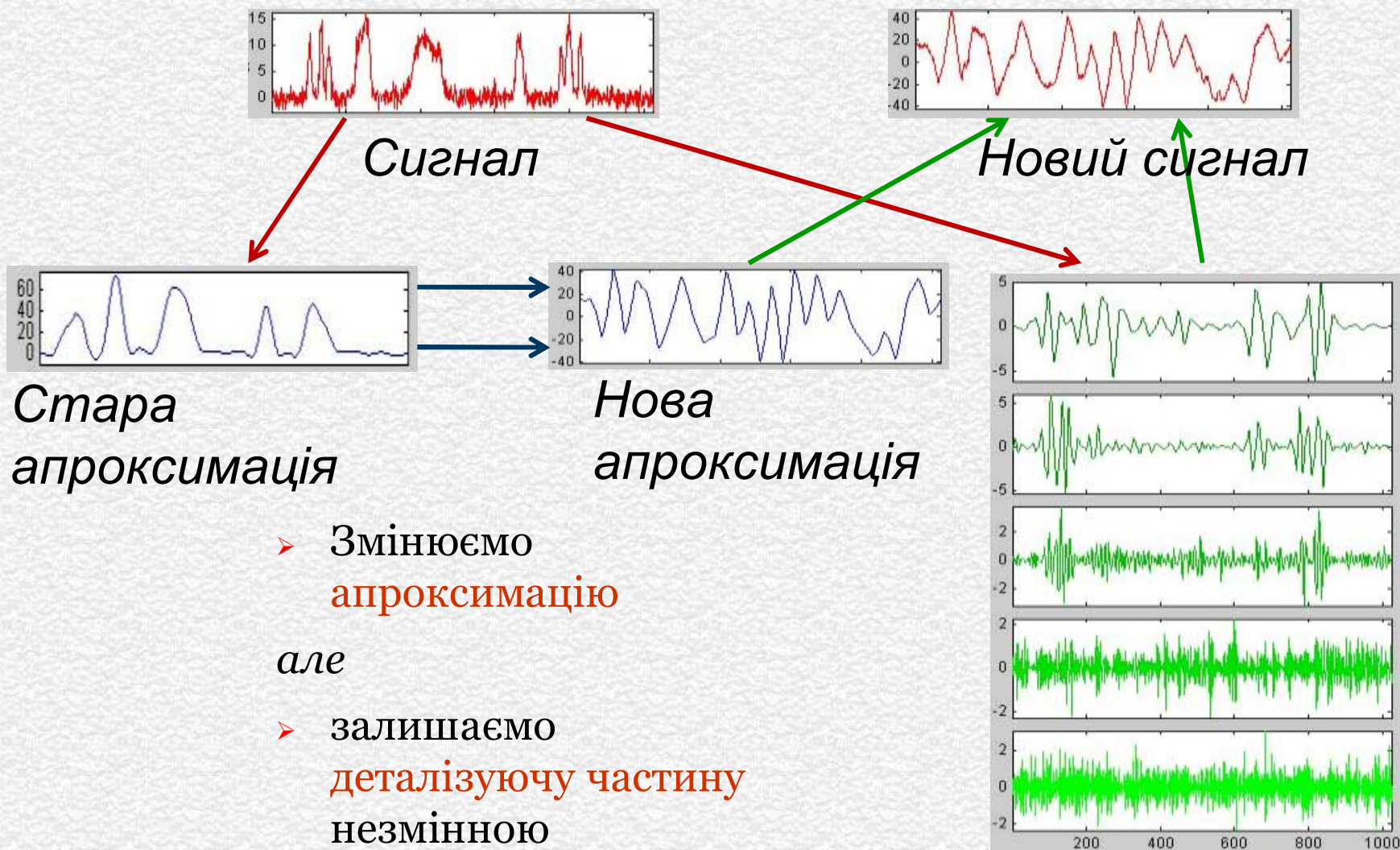
(А.А.Давидов, перший віце-президент, російського товариства соціологів)



Швидке зростання кримінальних злочинів в СРСР мало місце в 1984 і 1990.

Це непомітно на початковому сигналі.

Забезпечення групової анонімності



Загальна схема забезпечення групової анонімності

1. Побудова знеособленого мікрофайлу
2. Визначення захищаємих комбінацій сутнісних V_i та параметризуючого P_i атрибутів, $i=1, \dots, k$
3. Цикл по i від 1 до k :
 - a) вибір типу задачі (кількісна, концентраційна чи різницева)
 - b) вибір мети задачі (перенесення існуючих екстремумів, додавання хибних екстремумів, збереження мат.очікування та СКВ, ...)
 - c) розв'язання задачі (вейвлет-перетворення, нормалізація)
 - d) модифікація мікрофайлу шляхом обміну значеннями респондентів

метрика $InfM(r, r^*) = \sum_{p=1}^{n_{ord}} \alpha_p \left(\frac{r(I_p) - r^*(I_p)}{r(I_p) + r^*(I_p)} \right)^2 + \sum_{k=1}^{n_{nom}} \beta_k \left(\chi(r(J_k), r^*(J_k)) \right)^2$



Наукова новизна, 1

- ▼ *вперше виділено* для систем з обробки демографічної інформації **архітектурний рівень інтелектуального аналізу даних**, котрий забезпечує пошук прихованих закономірностей розподілу даних, тим самим розширюючи існуючі можливості з введення, структурованого обліку та аналітичної обробки демографічних даних;
- ▼ *вперше введено* **поняття групової анонімності даних** та *запропоновано* спеціальну метрику і стратегію застосування останньої для отримання оптимального маскуванню відповідних конфіденційних даних в мікрофайлі; на відміну від використовуємої наразі анонімізації даних про окремих респондентів **групова анонімність** дозволяє сформулювати та формалізувати задачу приховування інформації про особливості розподілу групи респондентів;
- ▼ *вперше поставлені* кількісна, концентраційна та концентраційно-різницева **задачі забезпечення групової анонімності даних** та *запропоновано* **метод їх розв'язання на основі застосування вейвлет-перетворень**;



Наукова новизна, 2

- ▼ *вперше розроблено метод пошуку* заздалегідь визначеного **шаблону** в цифровому сигналі, побудованому по статистичним даним, який, на відміну від інших методів, базується на використанні **поліномів наближення в просторі з породжуючим елементом**;
- ▼ *удосконалена відома загальна модель статистичних бізнес-процесів* шляхом введення нових процесів “**знеособлення**” і “**забезпечення конфіденційності**” та виокремлення процесу **пошуку** — як в інформаційно-обліковій статистичній системі (за допомогою нерегламентованих запитів) чи в інформаційно-аналітичній статистичній системі (через аналітичні запити), так і при застосуванні інтелектуального аналізу даних для автоматизованого пошуку в них певних закономірностей, що дозволяє надавати більш повний та точний опис процесів обробки демографічної інформації;
- ▼ *удосконалена модель Захмана* функціонально-інформаційної архітектури за рахунок її **розширення рівнем системної архітектури**, який відповідає, зокрема, й за інтелектуальний аналіз даних; дана модель була конкретизована для статистичної галузі України, що дозволяє її застосувати при побудові національних систем обробки демографічної інформації.



Впровадження

- ▼ АС “Перепис-2001” – в Головному міжрегіональну управлінні статистики у м. Києві та у 26 регіональних статистичних управліннях України
- ▼ АС “Перепис-2001 Аналітик” – в Головному міжрегіональну управлінні статистики у м. Києві
- ▼ АС “Перепис-Молдова 2004” – в Національному Бюро Статистики Республіки Молдова
- ▼ Техпроект на Комплекс обробки даних Всеукраїнського перепису населення 2012 р.

