

# РОБОТА З ВЕЛИКИМИ ДАНИМИ - ПОКАЗНИКАМИ СОЦІО-ЕКОЛОГО-ЕКОНОМІЧНОГО РОЗВИТКУ РЕГІОНУ

**Н. Б. Шаховська**

Доктор технічних наук, доцент,  
декан базової підготовки

Інститут комп'ютерних наук та інформаційних технологій\*

E-mail: natalya233@gmail.com

**Ю. Я. Болюбаш**

Здобувач\*

E-mail: bol\_jura@ukr.net

\*Кафедра інформаційних систем та мереж  
Національний університет "Львівська політехніка"  
вул. С. Бандери, 12, м. Львів, Україна, 79012

*У статті розглянуто особливості аналізу соціо-еколого-економічних даних та вказано, що їх доцільно опрацьовувати як Великі дані. Вказано особливості процесів консолідації даних. Визначено інформаційну технологію для роботи з даними регіону. Спроектовано схему сховища даних та обрано метод прогнозування значень показників регіону*

*Ключові слова: простір даних, великі дані, часовий ряд, геодані, слабо-структуровані дані, прогнозування, екосистема*

*В статье рассмотрены особенности анализа соціо-еколого-економических данных и указано, что их целесообразно обрабатывать как Большие данные. Указаны особенности процессов консолидации данных. Определена информационную технологию для работы с данными региона. Спроектировано схему хранения данных и выбран метод прогнозирования значений показателей региона*

*Ключевые слова: пространство данных, большие данные, временной ряд, геоданные, слабо-структурированные данные, прогнозирование, экосистема*

## 1. Вступ

Забезпечення екологічної безпеки неможливе без врахування існуючих тенденцій розвитку екосистем. Еколого-економічний прогноз є невід'ємною складовою частиною комплексного прогнозу соціально-економічного розвитку як країни в цілому, так і її окремих регіонів (районів) на перспективу і дозволяє оцінити характер стану природного середовища регіону при різних варіантах його розвитку, визначити домінуючі зв'язки між екологічною й іншою підсистемами. Для Західного регіону, зокрема Верхобужжя як історичного центру Золотого кільця екологічний прогноз має особливе значення, будучи домінуючим фактором при прийнятті рішень, які стосуються їх майбутнього господарського розвитку.

Останнім часом значення завчасного виявлення негативних тенденцій розвитку екосистем ще більш зросло, оскільки погіршення якості середовища стає істотно важливим фактором росту соціальної напруги.

Для комплексного аналізу регіону та прогнозування його розвитку необхідно:

1. Зберігати і керувати інформацією розміром у петабайти;

2. Опрацьовувати як структуровану, так і неструктуровану (у вигляді текстових звітів) інформацію, працювати з картографічними даними;

3. Аналізувати різномітну інформацію, використовуючи як консолідаційний, так і федеративний підхід.

Аналіз та моделювання соціально-економічних, регіональних систем необхідно виконувати з урахуванням таких характерних особливостей:

1. Регіон розглядається як складна слабоструктурована система, методологією дослідження якої є системний аналіз, необхідність використання якого викликана: наявністю великої кількості складних взаємопов'язаних причинно-наслідкових зв'язків між факторами, розглянутими в описі складної системи, результат дії яких не завжди очевидний при прийнятті рішень, необхідністю дослідження стохастичних систем в умовах невизначеності, неоднозначності;

2. Регіон – соціальна система, тому в ній домінують і враховуються природні і психологічні (пов'язані з інтересами людей та ін) чинники. При прийнятті рішень необхідно враховувати довгострокові інтереси суспільства. Рівень розвитку регіону покликаний, у першу чергу, забезпечувати умови відтворення людського життя;

3. Регіон – динамічна система. Необхідно вивчати динаміку розвитку системи, проводити аналіз процесів росту з урахуванням загального життєвого циклу регіону і його частин (населення, підприємства, житловий фонд та ін), адаптивної еволюції.

4. Регіон є саморегулюючою системою. Управління йде через внутрішньоорганізаційні процеси саморегулювання і полягає у зміні законів і методів внутрішнього управління.

На слабкість адміністративних заходів, неефективність цільового фінансування та інших адміністра-

тивних програм, не пов'язаних з приведенням у дію економічних регуляторів, вказував ще Форрестер [1]. Соціальні програми призводять до зрушень, порушень балансу;

5. Існує також конфлікт між цілями довгострокового планування і короткостроковими рішеннями;

6. Умовою нормального розвитку в системі є підтримання економічної рівноваги (балансу ресурсів у системі).

Методологічною основою моделювання соціально-економічного розвитку регіону є системний аналіз, центральною процедурою якого, як відомо, є побудова узагальненої (єдиної) моделі регіону, що відображає всі фактори і взаємозв'язки реальної системи.

На практиці це пов'язано з створенням комплексу моделей з розвиненими динамічними та інформаційними зв'язками між моделями усіх рівнів.

## 2. Аналіз останніх досліджень та постановка задачі

Проблема швидкого отримання різноманітних даних (сенсорних числових, текстових документів, графіків тощо) з метою формування на їх основі оперативних рішень постала ще у роки 2-ї світової війни і активно розвивається для застосування в атомних проєктах, управлінні ракетами, навігації, управлінні бойовими діями.

Опрацювання та аналіз таких різноманітних даних використовується для моделювання розвитку подій та ситуацій, а також в системах підтримки прийняття рішень.

Започаткували вивчення цієї проблеми фон Нейман, розробки компанії IBM, науковці школи Лебедева С. О. (спеціалізована ЕОМ), Глушкова В. М. (системний аналіз, теорія конфліктних ігор, проблемно-орієнтовані системи моделювання та опрацювання даних), що призвело до розвитку мов, блокового програмування, систем підтримки прийняття рішень.

Проте зміна класу досліджень – від оперативного до аналітичного, поява нових типів даних, необхідність швидкого доступу до них, зумовила збільшення інтересу до проблеми інтеграції та опрацювання даних з метою підвищення якості управлінських рішень. Найбільший пік активності дослідників у сфері інтеграції припадає на 90-ті рр. XX ст. та на наш час [2] у зв'язку з бурхливим розвитком Business Intelligence та збільшенням можливостей сховищ даних (збільшення обсягів збережених даних, наявність процедур аналітичного опрацювання даних – OLAP).

Особливістю сучасних досліджень є аналіз не лише типів даних (описів), але й семантики. Особливо активний розвиток засобів для оперативного збору різноманітних даних, завантаження їх у сховище даних, аналізу та прогнозування спостерігається в сферах енергетики та адміністративного керування, нафтогазовому секторі [2].

Схема отримання інформації органами керування регіоном передбачає створення статистичних звітів іншими об'єктами галузі за наперед визначеною формою з покровним агрегуванням інформації від одного об'єкту до іншого (рис. 1).

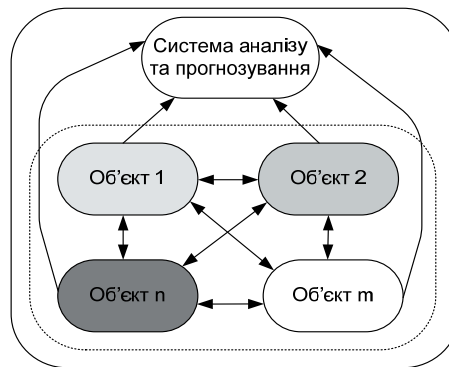


Рис. 1. Схема рівноправного обміну даними

Це приводить до того, що особа, яка отримує інформацію, бачить її лише в агреговану вигляді за жорстко визначеними критеріями групування, а деталізована інформація потрапляє зі значним запізненням.

Тому рішення, які можуть бути прийняті у такому випадку, недостатньо враховують усі особливості розвитку регіону.

Процес консолідації даних для аналізу та прогнозування розвитку регіону викликає задачі:

- підвищення оперативності отримання, аналізу та використання інформації, необхідної для підтримання прийняття рішень щодо керування регіоном;
- підвищення якості та дієвості керуючих рішень завдяки оперуванню достовірною інформацією, отриманою безпосередньо від джерела;
- визначення нових аспектів діяльності регіону завдяки аналізу даних, які не потрапляли у традиційні звіти, і тому не враховувалися при прийнятті рішень;
- своєчасного виявлення негативних тенденцій розвитку з метою їх подальшого усунення.

Інформаційний бум призвів до збільшення кількості даних, накопичених у багатьох предметних галузях у тисячі разів. До таких областей відноситься і державне управління. Кількість зібраної інформації зростає експоненційно. Так, за дослідженням IDC Digital Universe Study, проведеним на замовлення компанії EMC, сумарний обсяг світових даних у 2005 році складав 130 ексабайт, до 2011 року він зріс до 1227 EB, а за останній рік знову подвоївся, досягши 3 ZB (зетабайт). Прогноз, здійснений тим же дослідженням, показує, що до 2015 року обсяг цифрових даних зростає до 7.9 ZB. Розмір окремих баз даних зростає так само швидко і подолав петабайтний бар'єр. Більшість зібраних даних на даний час не аналізується, або ж проходить лише поверхневий аналіз [3].

## 3. Мета, завдання, об'єкти і методи досліджень

Основними проблемами, які виникають при обробці даних, є відсутність методів аналізу, придатних до застосування через їх різноманітність (для регіону – це і числові дані, і геодані, слабо структуровані звіти тощо), потреба у значних людських ресурсах для підтримки процесу аналізу даних, висока обчислювальна складність наявних алгоритмів аналізу

та стрімке зростання обсягу зібраних даних. Вони призводять до постійного зростання часу аналізу навіть при регулярному оновленні апаратних засобів серверів, а також – необхідність роботи із розподіленими базами даних, можливості яких більшість існуючих методів аналізу даних не використовують ефективно.

Таким чином, виникає задача розроблення ефективного методу аналізу даних, що може застосовуватись до розподілених баз даних різних предметних областей. Тому для регіону доцільно розробляти методи та засоби консолідації даних та використання їх для аналізу.

#### 4. Розроблення інформаційної технології опрацювання даних регіону

Великі дані (Big Data) в інформаційних технологіях – набір методів та засобів опрацювання структурованих і неструктурованих різнотипних динамічних даних великих обсягів з метою їх аналізу та використання для підтримки прийняття рішень. Є альтернативою традиційним системам управління базами даних і рішеннями класу Business Intelligence. До цього класу відносять засоби паралельного опрацювання даних (NoSQL, алгоритми MapReduce, Hadoop) [3, 4].

Визначальними характеристиками для Великих даних є обсяг (volume, в сенсі величини фізичного обсягу), швидкість (velocity в сенсах як швидкості приросту, так і необхідності високошвидкісної обробки та отримання результатів), різноманіття (variety, в сенсі можливості одночасної обробки різних типів структурованих і напівструктурованих даних).

Однією з технологій, що доцільно використовувати для роботи з Великими даними регіону, є простір даних.

Простір даних – це блоковий вектор, що містить множину інформаційних продуктів предметної області, поділену на три блоки: структуровані дані (бази, сховища даних), напівструктуровані дані (XML, електронні таблиці) та неструктуровані дані (текст). Над цим вектором та його окремими елементами визначено операції та предикати, які забезпечують [5, 7 – 10]:

- перетворення різних елементів вектора один в одного;
- об'єднання елементів одного типу;
- пошук в елементах за ключовим словом.

Розроблення еколого-економічного прогнозу регіону зводиться до вирішення трьох конкретних завдань:

- соціально-економічна оцінка стану природного середовища в регіоні в даний час і перспективі, розроблення на її основі системи заходів по повному запобіганню чи максимальному пом'якшенню негативного впливу господарської діяльності на навколишнє середовище;

- визначення й врахування можливих наслідків змін у природному середовищі в результаті господарської діяльності і техногенних процесів, їх вплив на спеціалізацію і комплексний розвиток господарства регіону;

- врахування прогнозів еколого-економічних процесів у контексті загального комплексного прогнозу соціально-економічного розвитку регіону шляхом формування ряду критеріїв і обмежень як по ресурсах, так і за допомогою показників якісного стану навколишнього середовища.

Тоді задачею прогнозування є знаходження залежності між датою, параметрами ресурсів (оздоровчий, історичний, культурний, відпочинковий), отриманих з бази даних вказаної структури, на основі аналізу попередньо отриманих даних.

Динаміка зміни виробленої потужності може бути охарактеризована стосовно якогось базисного (зазвичай першого) спостереження і величиною зміни сусідніх рівнів. У якості статистичних характеристик часового ряду  $Y_i$ ,  $i=1..n$  використовуються наступні величини: середнє арифметичне  $\bar{Y} = \frac{1}{N} \sum_{j=1}^N Y_j$ , середній абсолютний приріст  $\bar{Y} = (Y_n - Y_1)/(N-1)$ , де  $N$  – число рівнів ряду,  $Y_i$  – рівні ряду.

Відповідно до методу перевірки істинності різниці середніх початковий часовий ряд розбивається на дві однакові (або майже однакові) частини, після чого перевіряється гіпотеза про істотність різниці середніх для цих частин.

Недолік методу полягає в неможливості правильно визначити наявність тренда у тому випадку, коли часовий ряд містить точку зміни тенденції в районі середини ряду.

У методі Форстера-Стюарта гіпотеза про відсутність тренда перевіряється за допомогою допоміжних функцій [1]:

$$L = \sum_{t=2}^N I_t, \quad I_t = U_t - V_t, \quad u_t = \begin{cases} 1, & Y_t < Y_{t-1}, \dots, Y_1 \\ 0 & \end{cases}$$

$$v_t = \begin{cases} 1, & Y_t > Y_{t-1}, \dots, Y_1 \\ 0 & \end{cases}$$

Гіпотеза про відсутність тенденції не приймається, якщо розрахункове  $t$ -значення більше табличного на вибраному рівні значенності 0.95.

Перевірка однорідності даних здійснено на основі критерію Ірвіна, що заснований на порівнянні сусідніх значень ряду.

Відповідно до нього розраховується характеристика

$$t: \quad t = \frac{Y_t - Y_{t-1}}{Y}$$

Отримані значення порівнюються з табличними значеннями. Оцінка властивостей зводиться до дослідження автокореляційної функції вихідного і різницевого рядів.

Аналіз автокореляції виконується за допомогою графіка і критичних значень коефіцієнтів, встановлених експертом [6].

Далі спроектуємо структуру сховища даних для збереження показників розвитку регіону і його елементів (рис. 2).

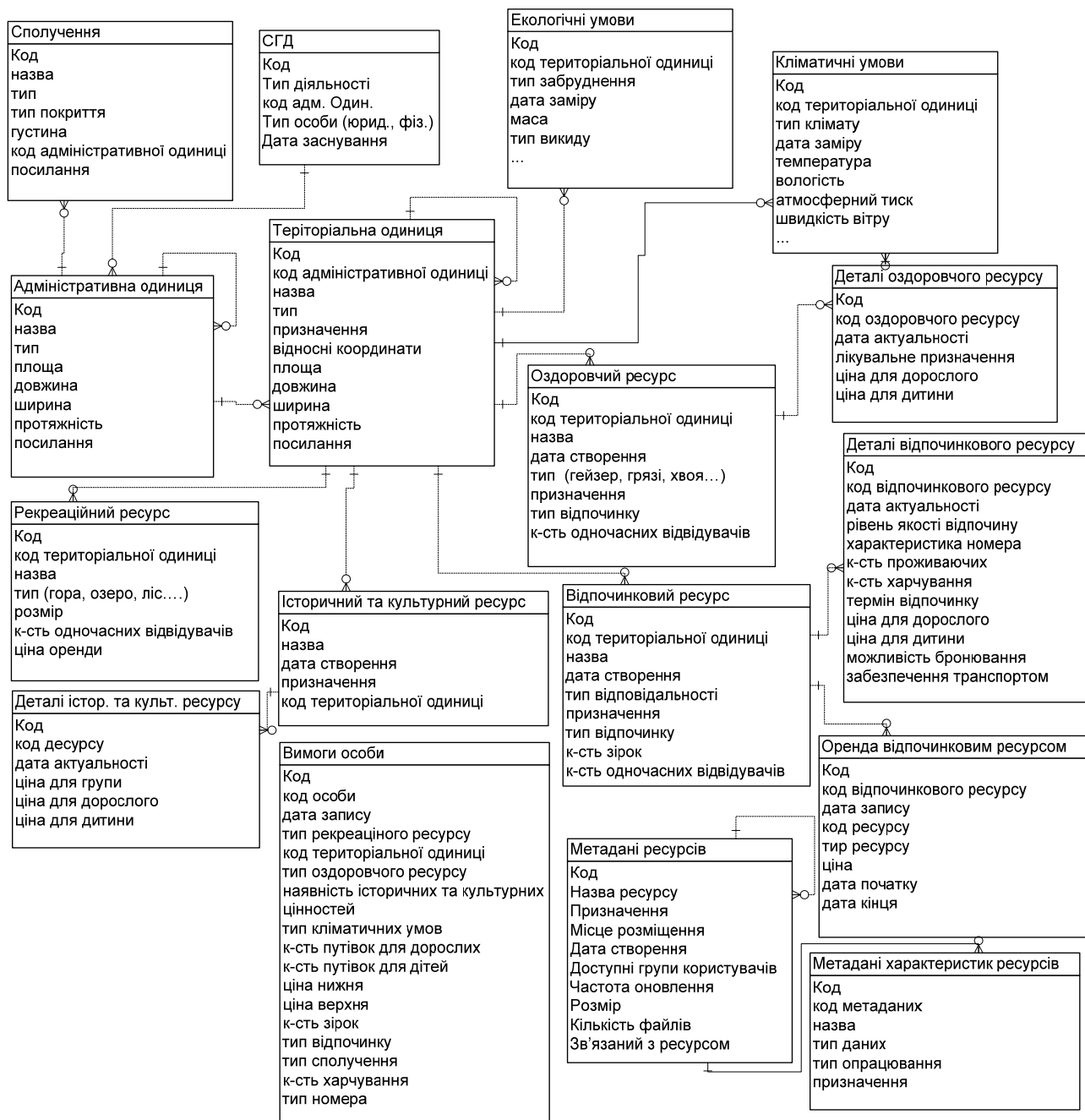


Рис. 2. Логічна модель даних регіону

## 5. Висновки

Таким чином, прогнозування розвитку регіону здійснюється на основі часових рядів з використанням аналізу попередніх станів регіону. Найдоцільнішою технологією для роботи з Великими даними регіону є простір даних.

Наукова новизна полягає у застосуванні технології простору даних до опрацювання Великих даних, а також розробленні логічної моделі даних регіону.

Практична цінність роботи полягає у використанні методу часових рядів до прогнозування розвитку регіону за узагальненими показниками.

## Література

1. Горчаков, А. Математичний апарат для інвестора. Аналіз та прогнозування часових рядів [Текст] / А. Горчаков // Фінансовий ринок України. - 2007. - 12. - С. 38-42.
2. Гриценко, В. И. Информационные технологии: тенденции, пути развития / В.И.Гриценко, А.А.Урсаев // Управляющие системы и машины. - 2001, № 5. - С.3-20.

3. Hilbert, M. Big Data for Development: From Information- to Knowledge Societies" // SSRN Scholarly Paper No. ID 2205145). Rochester, NY: Social Science Research Network. – 2013. [Електронний ресурс]. - [Режим доступу]: <http://papers.ssrn.com/abstract=2205145>.
4. White T. Hadoop: The Definitive Guide. O'Reilly Media. // ISBN 978-1-4493-3877-0. - 2012. – P. 3.
5. Шаховська, Н. Б. Формальне подання простору даних у вигляді алгебраїчної системи / Н.Б. Шаховська // Системні дослідження та інформаційні технології = System research & information technologies: міжнародний науково-технічний журнал / Національна академія наук України, Інститут прикладного системного аналізу. – Київ, 2011. – № 2. – С.128 – 140.
6. Кісь, Я. П. Інтелектуальні геоінформаційні системи. Міжнародний досвід та шляхи розвитку в Україні / Я.П.Кісь, Н.Б.Шаховська, О.Б.Вальчук. // Інформаційні системи та мережі №653. - Л. : Вид-во Нац. ун-ту «Львів. політехніка», 2008. - С. 139-145.
7. Шаховська, Н. Б. Методи опрацювання консолідованих даних за допомогою просторів даних / Н.Б. Шаховська // Проблеми програмування / Національна академія наук України, Інститут програмних систем НАН України, 2011, № 4. – С. 72-84.
8. Шаховська, Н. Б. Особливості інтеграції даних інформаційних систем Національного університету «Львівська політехніка» / Н.Б. Шаховська, Д.О.Тарасов // Складні системи і процеси / Запорізький інститут державного та муніципального управління. – 2009.- №2. – С. 98 – 109.
9. Шаховська, Н. Б. Алгебраїчна система класу „простір даних» / Н.Б. Шаховська // Науковий вісник Чернівецького університету. Серія Фізика. Електроніка: Збірник наукових праць. – Чернівці, 2009. – № 479: Тематичний випуск «Комп'ютерні системи та компоненти». – С. 48 – 51.
10. Калининченко, Л. А. Методы синтеза канонических моделей, предназначенных для достижения семантической интероперабельности неоднородных источников информации / Л. А. Калининченко, С. А. Ступников, Н. А. Земцов – Москва: ИПИ РАН, 2005.-84 с.

*Розроблено метод кодування символів у когерентному оптичному каналі на основі фазової модуляції, який розширює діапазон модуляції фази. Кодування символів здійснюється ортогональними гармоніками що утворюються шляхом модуляції фази. Кожна гармоніка спектру кодує один розряд символу у трійчному коді, що дозволяє передавати спеціальні символи для управління цифровим потоком в оптичному каналі*

*Ключові слова: когерентний оптичний канал, фазова модуляція, кодування символів, ортогональний спектр радіочастот*

*Разработан метод кодирования символов в когерентном оптическом канале на основе фазовой модуляции, который расширяет диапазон модуляции фазы. Кодирования символов осуществляется ортогональными гармониками, которые формируются путем модуляции фазы. Каждая гармоника спектра кодирует один разряд символа в тричном коде, что позволяет передавать специальные символы для управления цифровым потоком в оптическом канале*

*Ключевые слова: когерентный оптический канал, фазовая модуляция, кодирование символов, ортогональный спектр радиочастот*

УДК 621.391

# МЕТОД КОДУВАННЯ СИМВОЛІВ У КОГЕРЕНТНОМУ ОПТИЧНОМУ КАНАЛІ ІНТЕГРОВАНОЇ ТЕХНОЛОГІЇ ТЕЛЕКОМУНІКАЦІЙ

В. І. Тіхонов

Кандидат технічних наук, доцент  
Кафедра «Мережі зв'язку»Одеська національна академія зв'язку  
ім. О.С. Попова

вул. Ковальська, 1, м. Одеса, Україна, 65029

E-mail: victor.tykhonov@onat.edu.ua

## 1. Вступ

Однією з перспективних технологій фізичного рівня в телекомунікаціях є когерентні оптичні комунікації (Coherent Optical Communications – COC), які уможливають вторинне ущільнення оптичних ліній зв'язку в технологіях xWDM (WDM, CWDM, DWDM

тощо) [1, 2]. Методи вторинного ущільнення оптичних каналів можна розділити на дві основні категорії: радіочастотне ущільнення (Radio Frequency Optical – RFO) і оптичне вторинне ущільнення – (Orthogonal Wave Division Multiplexing (OWDM) [3, 4]. Найбільш відомими методами RFO-ущільнення є ортогональне радіочастотне ущільнення оптичного каналу (RFO-OFDM), а