

## РОЗДІЛ 27

**МЕТОДЫ ПОВЫШЕНИЯ КАЧЕСТВА КЛАССИФИКАЦИИ  
СИМВОЛОВ ПРИ ОБРАБОТКЕ ТЕКСТОВЫХ  
ИЗОБРАЖЕНИЙ**

*Аннотация.* В работе рассматриваются методы классификации, применяемые при сжатии файла с битональным изображением текста, полученным сканированием или цифровым фотографированием. Особое внимание обращается на используемые при этом меры различия двух изображений символов, выделенных из изображения текста. Эти меры различия позволяют с той или иной степенью уверенности считать символы на сравниваемых изображениях или совпадающими, или различными. Для известных на сегодняшний день алгоритмов классификации, включая хорошо известный алгоритм JB2, приведены количественные характеристики классификации – число классов, получаемых этими алгоритмами для изображения стандартной страницы текста. Чем меньше это число, тем качество классификации считается выше, так как дает лучшее сжатие файла с изображением текста. Рассмотрены также методы ускорения алгоритмов, классифицирующих изображения символов, и повышения удобочитаемости восстановленного после сжатия изображения текста.

*Ключевые слова:* изображение текста, методы классификации, сжатие данных.

*Abstract.* Methods are in-process examined classifications, applied at the compression of file with the bi-tonal image of photo-typograph got scan-out or digital photographing. The special attention applies on the in-use here measures of distinction of two images of characters, abstracted from the image of text. These measures of distinction allow with one or another degree of confidence to count characters on the compared images or consilient, or different. For the algorithms of classification known to date, including the known algorithm of JB2 well, quantitative descriptions of classification – number of classes, got these algorithms for the image of standard page of text are resulted. What less than it is a number, quality of classification is considered that higher, because gives the best compression of file with the image of text. The methods of acceleration of algorithms, classifying the images of characters, and increases of easy-to-readness of the text recovered after the compression of image are considered also.

*Keywords:* image of text, methods of classification, compression of data.

**Введение и постановка задачи.** Высокие результаты, которые демонстрирует алгоритмом JB2 (формат DjVu), объясняются тем, что он использует классификацию символов. Вообще идея сжатия информации с

помощью классификации очень проста и идеально подходит для сжатия изображений текста.

Пусть необходимо сжать некую информацию, которую можно разбить каким-то образом на элементы. Если эти элементы информации объединить в классы так, чтобы в каждом классе находились тождественные (pattern matching) или почти тождественные (soft pattern matching) элементы, то нет нужды хранить все элементы информации – достаточно хранить только по одному элементу из каждого класса. Совокупность таких элементов – представителей классов – называется словарем. Кроме того для восстановления информации нужно еще иметь таблицу, называемую «картой размещения классов», которая для каждого класса указывает, в каком месте исходной информации находятся его элементы.

Ясно, что степень сжатия информации с помощью классификации тем выше, чем меньше классов образуется при классификации и чем больше элементов в каждом классе.

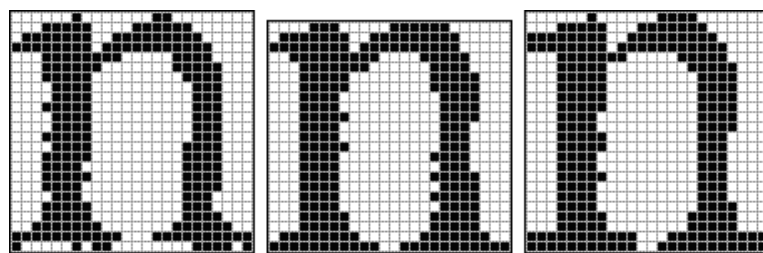
В случае сжатия изображения бинарного (далее черно-белого) текста естественным элементом информации является изображение отдельного символа (буквы, цифры, знака препинания и т.п.). Выделение символов не представляет собой особо трудную задачу. Во всех известных алгоритмах, включая алгоритм JB2, символы выделяются как связные области, состоящие из черных точек.

Следует заметить, что при этом некоторые грамматические символы распадаются на части (например, буква “ё” дает три символа), а некоторые (например, сочетания вида “fh”) объединяются в один. Кроме того метод непригоден для текстов с псевдо рукописным шрифтом. Сжатие таких текстов алгоритмом JB2 и другими катастрофически низкое.

Однако не это представляет собой главную трудность при классификации уже разделенных символов.

Цель настоящей статьи – дать обзор идей и методов, на которых основаны эти алгоритмы, и провести сравнение достигнутых с их помощью результатов по сжатию изображений текста.

На рис. 27.1, взятом из работы [32], представлены три случайно выбранные изображения буквы “n” из различных 257, входящих в изображение страницы текста формата A4, при разрешении сканирования 300 dpi.



*Рис. 27.1. Влияние шумов печати и сканирования на изображение символа «п»*

Легко верится, и это действительно так, что на странице не найдется ни одной пары символов “п”, полностью совпадающих друг с другом. То же, за редким исключением, относится и к другим символам, даже точкам. Причиной этого явления являются шумы (то есть случайные искажения), возникающие при печати страницы и ее последующем сканировании. Шумы печати в основном вызваны диффузией краски, жидкой или твердой, вдоль хаотически расположенных капилляров бумаги. Шумы сканирования – несовпадением контуров символа с матрицей сканера, подобно тому, как прямая наклонная линия на экране монитора отображается “ступеньками”.

Человеку легко заметить, что все три изображения, приведенные на рис. 27.1, представляют собой букву “п”. Однако пока не существует алгоритма, который мог бы установить тождественность этих символов с той же достоверностью, что и человек. Это и есть главная трудность, не позволяющая разбить изображения символов на классы, так чтобы одновременно выполнялись два условия:

**Условие 1.** В каждом классе находятся изображения только одного и того же символа;

**Условие 2.** Все изображения какого-либо символа находятся в одном классе.

Все алгоритмы классификации являются тем или иным компромиссом между этими условиями, причем условие 1 должно выполняться достаточно жестко, иначе в восстановленном тексте будут перепутаны символы, чем иногда грешит алгоритм JВ2. Например, иногда путает между собой буквы «b» и «h».

Соблюдение условия 1 влечет за собой ужесточение алгоритма сравнения изображений символов, так что условие 2, практически, невыполнимо. Это приводит к появлению значительно большего числа классов, чем количество символов, изображенных на странице, так как практически все символы дают по несколько классов своих изображений. Чем больше при классификации

образуется классов, тем больше словарь и (логарифмически) больше карта размещения классов. Как следствие, понижается степень сжатия. И хотя алгоритм JB2 и другие используют те или иные методы дополнительного сжатия словаря и карты, эффективность алгоритма в целом определяется качеством классификации, то есть количеством получившихся классов, которое в идеале (условие 2) должно совпадать с количеством символов, присутствующих в тексте, чье изображение сжимается.

Таблица 27.1

**Количество классов при рассматриваемых методах классификации**

Разрешение сканирования (dpi)	Количество классов в исходном изображении	Количество классов после основной классификации $\varepsilon_{opt} = 6\%$	Количество классов после второй классификации $\varepsilon_{opt} = 6\%$	Количество классов после классификации алгоритмом JB2
600 dpi	3558	197	72	314
500 dpi	3557	137	72	259
400 dpi	3557	130	71	199
300 dpi	3545	122	95	235
200 dpi	3890	237	148	451

В табл. 27.1 из работы [32] для различных разрешений сканирования показано количество классов, полученных предлагаемым алгоритмом классификации (предпоследний столбец) и алгоритмом JB2 (последний столбец). Первый столбец показывает разрешение, использованное при сканировании одной и той же страницы формата А4 с черно-белым текстом, набранным шрифтом Times New Roman, 12 pt. Второй столбец – количество классов при тождественной классификации, то есть классов, состоящих из полностью совпадающих изображений символов.

Из таблицы следует несомненное превосходство алгоритма ИЛЛ – будем для краткости называть так алгоритм, предложенный в упомянутой работе [32], – над алгоритмом JB2. Словарь ИЛЛ получается почти в три раза короче, чем словарь JB2. Отсюда вытекает и превосходство в степени сжатия той же страницы, что показывает следующая таблица, относящаяся к той же странице, что и табл. 27.1.

Не слишком существенное различие между коэффициентами сжатия, продемонстрированными алгоритмами ИЛЛ и JB2, объясняется тем, что авторы

алгоритма ИЛЛ интересовались только классификацией выделенных изображений символов и не оптимизировали алгоритм дополнительного сжатия словаря и карты размещения классов (использовался универсальный алгоритм без потерь 7z).

Кроме того табл. 27.2 показывает, что применение лучшего для сжатия размытых изображений алгоритма JPEG 2000 мало что дает при сжатии изображения черно-белого текста без иллюстраций.

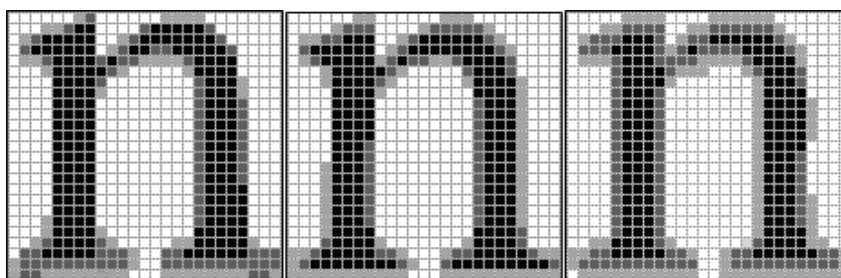
Таблица 27.2

**Сравнительная степень сжатия изображения текста**

Разрешение сканирования (dpi)	200	300	400	500	600
Исходный размер файла (kb)	505,3	1080,2	2003,9	3111,2	4498,0
<b>Методы</b>	<b>Коэффициент сжатия</b>				
JPEG 2000	3,8	3,74	3,76	3,75	3,75
JBIG2	8,2	11,2	16,7	20,9	25,1
JB2	52,6	124,1	202,4	272,9	330,7
ИЛЛ	62,3	135,0	250,4	353,5	436,7

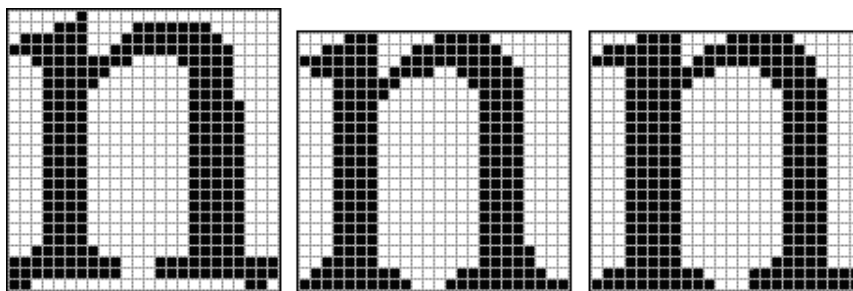
**Дополнительные меры повышения качества классификации символов.**

Своей высокой эффективностью алгоритм ИЛЛ обязан не только удачно выбранной и легко вычисляемой мере отличия [32], но и в большей степени тому, что он использует процедуру статистического усреднения. Это означает, что после проведения классификации для каждого класса находится усредненное изображение. Процедура усреднения состоит в наложении друг на друга всех изображений класса, совмещающем их “центры тяжести”, и вычисления среднего значения яркости для каждой точки с последующим округлением. На рис. 27.2 иллюстрируется процесс усреднения для трех классов изображений одного и того же символа “п”. Черные точки означают, что среднее значение яркости в них равно 0, серые – что среднее значение меньше или равно 0,5, а светлые – больше 0,5.



*Рис. 27.2. Суммарные изображения трех различных классов символа «п»*

При округлении черные и серые точки превращаются в черные, а светлые – в белые. Результат показан на рис. 27.3.



*Рис. 27.3. Изображения средних символов в трех классах символа «n»*

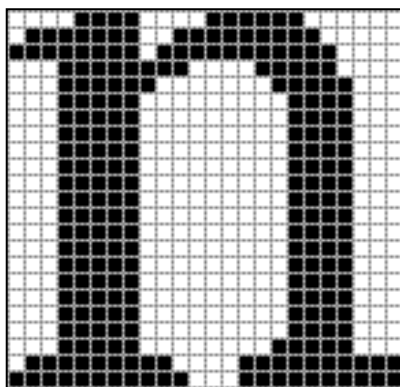
Совокупность полученных средних изображений в алгоритме ИЛЛ используется как промежуточный словарь, то есть все исходные изображения символов более не используются. Легко видеть (см. рис. 27.3), что благодаря процедуре усреднения искажения, вызванные шумами печати и сканирования, заметно меньше, чем в исходных изображениях (см. рис. 27.1).

Однако в полученном промежуточном словаре усредненных изображений по-прежнему присутствуют разные изображения одних и тех же символов, хотя и в значительно меньшем количестве, чем до классификации. Их количество для тестовой страницы указано в третьем столбце табл. 27.1.

Качество классификации можно улучшить, применив еще одну классификацию, призванную объединить в один класс изображения одного и того же символа. Для этой классификации, как и для предыдущей, используется нечувствительная к контурным шумам мера отличия  $\mathcal{E}$  с тем же порогом  $\mathcal{E}_{opt}$ , но алгоритм классификации выбирается другим.

Дело в том, что теперь разные изображения одного того же символа очень близки друг к другу – случайная компонента изображений в значительной мере подавлена. Поэтому существенно меньше опасность спутать изображения двух разных символов.

После получения новых классов, изображения в каждом из них снова усредняются, и получившийся набор изображений представляет собой окончательный словарь. На рис. 27.4 показано теперь единственное изображение символа “n”, вошедшее в словарь. Эффективность повторной классификации можно проследить, сравнивая третий и четвертый столбцы табл. 27.1.



*Рис. 27.4. Представитель класса изображения символа «п», который попадает в словарь символов*

Сравнивая рис. 27.4 с рис. 27.1, можно утверждать, что контурные шумы практически отсутствуют и внешний вид символа значительно улучшился, по сравнению с вариантами его изображения в исходном тексте. Таким образом, качество восстановленного изображения текста в алгоритме ИЛЛ значительно выше исходного. В принципе так и должно быть, потому что значительная часть избыточной информация, которую устраняют алгоритмы сжатия изображений текста, состоит из случайных шумов печати и сканирования.

#### **Методы ускорения классификации символов**

В алгоритме ИЛЛ предварительная классификация проводится методом «просеивания», также с помощью очень быстро вычисляемой дополнительной меры различия. Изображению каждого символа сопоставляется трехмерный вектор, и мера различия между двумя изображениями символов – расстояние между соответствующими векторами, нормированными так, чтобы порог не зависел от размера шрифта и разрешения сканирования.

Вектор параметров каждого изображения символа, на сравнении которых проводится классификация, состоит из следующих величин:  $H$  – высота символа, то есть высота прямоугольника, в который вписано изображение символа;  $W$  – аналогично определенная ширина символа;  $P$  – периметр изображения символа, подсчитанный в пикселях вдоль внешних и внутренних его границ.

При сравнении размеров двух символов вычисляется разность их физических размеров, например, для высоты – величина

$$\Delta H = |H_1 - H_2| \frac{100}{res(dpi)},$$

где  $H_1$ ,  $H_2$  – высоты первого и второго сравниваемых символов в пикселях, и  $ras(dpi)$  – разрешение сканирования, измеряемое в точках на дюйм.

Таким образом, разность высот выражается в сотых частях дюйма. Аналогично вычисляется разность ширин  $\Delta W$  сравниваемых символов.

Переход к физическим линейным размерам оправдан тем, что на отклонения высоты и ширины изображений одного и того же символа в основном влияют шумы печати, а не сканирования. Последние могут изменить линейный размер не более, чем на 1 пиксель. Экспериментально установлено, что отклонение в линейных размерах, вызванное расплыванием краски, находится в пределах 0.01 дюйма. Поэтому линейные размеры символов вне зависимости от разрешения и размера шрифта можно считать значительно отличающимися, если  $\Delta H > 1$  и  $\Delta W > 1$ .

Длина периметра, определяемая как число граничных пикселей в изображении символа, мало искажается шумами печати и сканирования. В основном она выражает индивидуальные особенности символов, то есть гарантированно различает, например, такие буквы, как “n” и “r”. Поэтому для их сравнения вводится в рассмотрение безразмерная, то есть не зависящая ни от разрешения сканирования, ни от размера шрифта величина

$$\Delta P = \frac{|P_1 - P_2|}{\sqrt{P_1 P_2}} 100\%$$

где  $P_1$  и  $P_2$  – периметры сравниваемых символов.

Периметры символов считаются далекими, если  $\Delta P > 10\%$ . Если этому сопутствует значительное различие в линейных размерах:  $\Delta H > 1$  и  $\Delta W > 1$ , то можно с уверенностью утверждать, что сравниваемые изображения относятся к разным символам. Такая классификация, с одной стороны, имеет небольшое число классов, каждый из которых содержит изображения близких по начертанию символов. С другой стороны, выполнено условие 2, то есть все изображения одного и того же символа попадают в один и тот же класс.

#### **Дополнительное сжатие словаря и карты размещения классов.**

Размер графического файла, который сжат каким-либо алгоритмом, основанным на классификации, грубо говоря, состоит из размера словаря, и размера соответствующей карты размещения классов. Этот размер можно уменьшить, если дополнительно сжать каким-либо стандартным алгоритмом сжатия без потерь, например, алгоритмами RLE, Хаффмана, LZW и другими [3; 97]. Или их комбинациями. И хотя при создании стандартного формата



сжатия изображений текста на эти вопросы требуется обратить самое серьезное внимание, здесь мы на этом останавливаться не будем. Отметим только, что размер словаря во много раз превышает размер отвечающей ему карты (для одной страницы текста) и представляет собой набор изображений символов, для сжатия которых методы компрессии без потерь не являются самыми эффективными.

Рассмотрим предложенный в работе [33] алгоритм – будем называть его алгоритмом ИЛЛ2 – который сжимает словарь, рассматривая его не как последовательность бит, а как изображение, составленное из отдельных изображений входящих в него символов. Идея этого алгоритма состоит в следующем.

Если представить себе строку в изображении текста как прямоугольник с горизонтальным основанием, охватывающий строку, то вертикальным элементом строки называется пересечение этого прямоугольника с любой вертикальной линией шириной в один пиксель. На рис. 27.5 из работы [33] показано разбиение изображения буквы “е” на вертикальные элементы строки.



*Рис. 27.5. Изображение символа «е» и составляющие его вертикальные элементы строки*

Таким образом, все символы в строке, включая пробелы, можно представить как объединение ее вертикальных элементов одного и того же размера.

Алгоритм ИЛЛ2 рассматривает словарь как изображение строки, составленной из изображений всех входящих в него символов. (Фрагмент такого словаря показан на рис. 27.6.). Эта строка разбивается на вертикальные элементы, совокупность которых классифицируется отдельным способом. В результате получается новый словарь, состоящий из вертикальных элементов строки, и карта размещения соответствующих ему классов. Смысл этих процедур заключается в том, что новый словарь занимает намного меньше

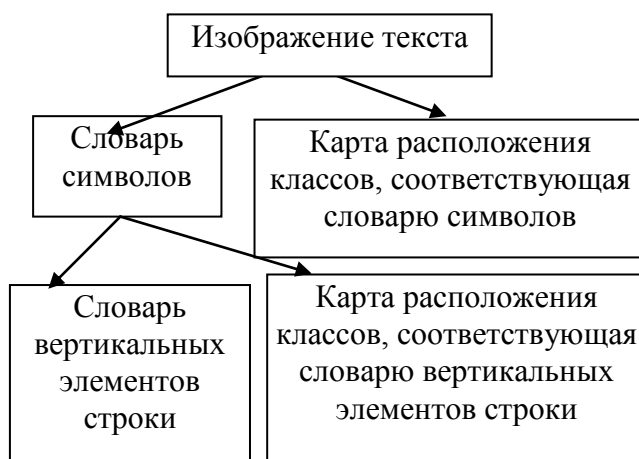
места, чем старый, а дополнительная карта размещения классов невелика по размеру.



**Рис. 27.6.** Изображение фрагмента словаря для изображения текста 300 dpi: *вверху – словарь до сжатия; внизу – восстановленный словарь*

Внизу, на рис. 27.6 показан словарь символов, восстановленный с помощью словаря вертикальных элементов строки и соответствующей ему карты расположения классов. Некоторое незначительное ухудшение качества заметно только при большом увеличении, которое используется на этом рисунке.

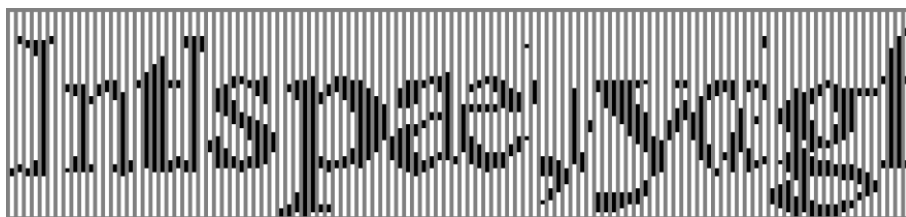
Таким образом, алгоритм ИЛЛ2 работает в два этапа, как это схематически изображено на рис. 27.7. На первом этапе (переход с первого уровня на второй) он совпадает с алгоритмом ИЛЛ, и отличается от последнего наличием еще одного этапа (переходом со второго уровня на третий), на котором сжимается словарь, полученный на первом этапе.



**Рис. 27.7.** Схема двухэтапной обработки изображения текста алгоритмом ИЛЛ2

Фрагмент словаря вертикальных элементов строки для изображения текста с разрешением 300 dpi представлен на рис. 27.8 (серые полосы введены для

наглядности как разделители между представителями отдельных классов вертикальных элементов строки).



*Рис. 27.8. Фрагмент словаря вертикальных элементов строки, полученный после автоматической классификации*

Перейдем к описанию алгоритма классификации вертикальных элементов строки, который базируется на автоматической классификации (алгоритме  $k$ -средних [3; 31]).

Вертикальные элементы строки, совокупность которых обозначим через  $x$ , рассматриваются как векторы с бинарными координатами (0 – черный цвет, 1 – белый). Единственным параметром классификации является  $k$  – число классов.

Некоторым образом из множества  $x$  выбирается элементов  $e_1, \dots, e_k$  – центров класса (в первом приближении). Все элементы изображения разбиваются на  $k$  классов  $S_1, S_2, \dots, S_k$  по правилу: для каждого элемента  $x$  находится ближайший в евклидовой метрике центр класса, после чего в один класс объединяются все элементы, имеющие один и тот же ближайший центр класса. Далее в каждом классе находятся новые центры класса (во втором приближении) путем усреднения элементов в каждом классе:

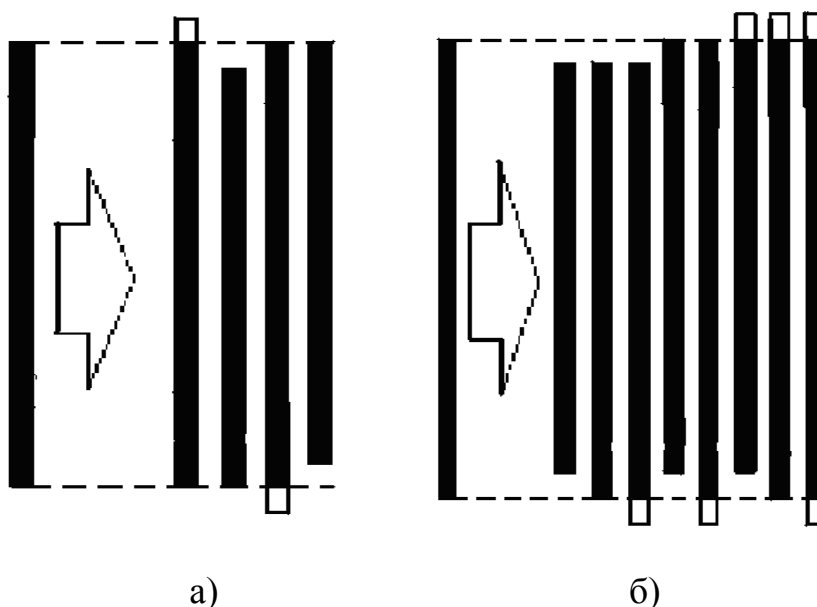
$$e_i = \frac{1}{N_i} \sum_{x \in S_i} x,$$

где  $N_i$  – число элементов в классе  $S_i$ , и  $i = 1, 2, \dots, k$ , и последующим округлением до 0 или 1).

После этого процедура повторяется, исходя из новых центров классов. Алгоритм заканчивает работу, когда центры классов перестают изменяться. Их совокупность – словарь вертикальных элементов строки.

Автоматическая классификация – очень быстрый алгоритм, дающий классификацию высокого качества, если правильно выбраны число классов  $k$  и набор центров первого приближения. В алгоритме ИЛЛ2 и то, и другое

определяется с помощью алгоритма просеивания, с мерой отличия, учитывающей контурный характер искажений, вносимых шумами печати и сканирования.



**Рис. 27.9. Возможные варианты близких связанных компонент вертикальных элементов строки: а)  $\varepsilon = 1$ ; б)  $\varepsilon = 2$**

Эта мера отличия  $\varepsilon$  двух вертикальных элементов строки равна бесконечности во всех случаях кроме тех, когда каждой связанной компоненте множества черных пикселей одного из них взаимно однозначно отвечает компонента другого так, что эти компоненты могут отличаться на один пиксель только на своих концах. Возможные варианты показаны на рис. 27.9. Во всех этих случаях мера отличия  $\varepsilon$  равна максимальному по каждой компоненте числу несовпадений. Иначе говоря,  $\varepsilon = 0$ , если вертикальные элементы строки полностью совпадают,  $\varepsilon = 1$ , если все пары отвечающих друг другу компонент совпадают кроме некоторой, отличающихся на один пиксель на одном из концов, верхнем или нижнем, и  $\varepsilon = 2$ , если найдутся пары компонент, отличающихся на один пиксель и на нижнем, и на верхнем концах. Во всех остальных случаях  $\varepsilon_{opt} = \infty$ .

Выбор такой меры отличия объясняется тем, что контурные шумы печати и сканирования могут, как правило, изменить черную компоненту вертикального элемента строки в ту или иную сторону лишь на один приграничный пиксель.

После разбиения всей совокупности вертикальных элементов на классы методом просеивания с порогом  $\varepsilon_{opt} = \varepsilon_{max} = 2$  в качестве центров нулевого приближения для алгоритма автоматической классификации выбираются средние в каждом из полученных классов за исключением тех классов, которые содержат 1–2 элемента.

Последнее объясняется тем, что шумы печати и сканирования порождают новые, как правило, многокомпонентные вертикальные элементы, смежные с вертикальными границами символов. А также сильно искажают соседние элементы, принадлежащие изображению символа. (См., например, рис. 27.1.) Такие элементы при просеивании дают классы, состоящие из одного-двух элементов (для стандартной страницы текста) и не выбираются в качестве центров класса. При автоматической классификации благодаря проводимому на каждой итерации усреднению, они не оказывают никакого влияния на конечный результат. Это обстоятельство оправдывает выбор автоматической классификации.

Отметим также, что предварительное применение алгоритма просеивания значительно улучшает сходимость автоматической классификации по сравнению со случайным выбором центров первого приближения и, следовательно, сокращает вычислительное время. Это объясняется тем, что уже на первой итерации центры нулевого приближения хорошо аппроксимируют члены своего класса с точностью, не меньшей, чем  $\sqrt{2n}$ , где  $n$  – число черных компонент.

Размер файла, который сжат каким-либо алгоритмом, основанным на классификации, в основном состоит из размера словаря и размера карты размещения классов. Причем при классификации выделенных изображений символов размер словаря весьма значителен. Поэтому предложенный в рассматриваемой работе алгоритм автоматической классификации вертикальных элементов строки, уменьшающий размер словаря, обеспечивает более высокую степень компрессии исходного изображения текста.

В табл. 27.3 для тестовой страницы А4, шрифт Times New Roman 12 pt представлены объемы словаря символов, словаря вертикальных элементов строки, а также общего объема словаря вертикальных элементов строки и соответствующей ему карты расположения классов. Параллельно приведены те же характеристики после дополнительного сжатия без потерь с использованием алгоритма 7z, который является модификацией словарного метода компрессии LZ77–LZMA.

Отметим, что полученный на втором этапе алгоритма ИЛЛ2 словарь вертикальных элементов строки в 20–30 раз меньше словаря символов, полученного на первом этапе. Однако после дополнительного сжатия результат не такой впечатляющий – в 2 – 4 раза (хотя это тоже очень хороший результат). Это объясняется тем, что словарь вертикальных элементов практически не содержит избыточную информацию, так что «дожимать» по сравнению с исходным словарем символов фактически нечего.

Карта расположения классов вертикальных элементов строки, по сути, аналогична обычному тексту, скажем, в ASCII кодировке, и при дополнительном сжатии уменьшается всего в 2–3 раза. Это в значительной мере уменьшает эффективность второго этапа алгоритма ИЛЛ2.

Таблица 27.3

### Объем словаря символов

Разрешение сканирования (dpi)	200	300	400	500	600
Размер словаря символов/ после дополнительного сжатия(kb)	34,8/4,4	46,4/5,0	67,7/4,7	94,6/5,4	140,2/6,7
Размер словаря вертикальных элементов строки / после дополнительного сжатия (kb)	1,2/1,0	2,5/1,5	4,4/2,1	7,3/3,0	10,8/3,8
Словарь вертикальных элементов строки + карта размещения их классов / после дополнительного сжатия (kb)	7,3/3,0	8,7/3,4	10,1/4,2	13,7/5,2	18,8/6,6

Таблица 27.4

### Выигрыш в сжатии изображения текста в результате компрессии словаря символов на втором этапе обработки

Разрешение изображения текста (dpi)	200	300	400	500	600
Коэффициент сжатия словаря символов (ИЛЛ = 1-й этап алгоритма ИЛЛ2)	7,9	9,28	14,4	17,51	20,92
Коэффициент сжатия словаря вертикальных элементов вместе с соответствующей ему картой размещения классов (II этап алгоритма ИЛЛ2)	11,6	13,64	16,11	18,19	21,24

Как показывает табл. 27.4, выигрыш в сжатии, полученный на втором этапе, существенен при разрешениях 200 и 300 dpi, что является неплохим результатом, так как эти разрешения наиболее популярны. При 400 dpi

выигрыш не столь заметен и практически отсутствует при больших разрешениях. Это объясняется тем, что с повышением разрешения увеличивается число вертикальных элементов строки на каждый символ, и карта расположения классов вертикальных элементов строки возрастает. Так как она сжимается слабо, то суммарный объем словаря вертикальных элементов строки и соответствующей ему карты с ростом разрешения сжимается все хуже и хуже.

Таблица 27.5

**Выигрыш в степени сжатия двухэтапного алгоритма в сравнении с алгоритмом JB2 (DjVu)**

Разрешение изображения текста (dpi)	200	300	400	500	600
Алгоритм ИЛЛ (или 1-й этап алгоритма ИЛЛ2)	62,38	135	250,48	353,54	436,7
Алгоритм ИЛЛ2	74,3	166,18	267,18	361,76	445,34
Преимуществов сжатии алгоритма ИЛЛ2 над алгоритмом ИЛЛ в (%)	<b>16%</b>	<b>19%</b>	<b>6%</b>	<b>3%</b>	<b>2%</b>
Алгоритм JB2 (DjVu)	52,63	124,16	202,41	272,91	330,73
Преимущество в сжатии ИЛЛ2 над алгоритмом JB2 в (%)	<b>29%</b>	<b>25%</b>	<b>24%</b>	<b>24,5%</b>	<b>25,5%</b>

В табл. 27.5 приведены значения коэффициентов сжатия изображения стандартной страницы текста алгоритмом ИЛЛ (или, что то же самое, после 1-го этапа алгоритма ИЛЛ2), алгоритмом ИЛЛ2 и алгоритмом JB2, а также преимущество в сжатии алгоритма ИЛЛ2 над алгоритмами ИЛЛ и JB2 в процентном соотношении.

Главный вывод, который можно сделать из этой таблицы, такой. Алгоритм ИЛЛ2 благодаря второму этапу повышает степень сжатия алгоритма ИЛЛ в основном при разрешениях сканирования в 200 и 300 dpi и, таким образом обеспечивает сжатие примерно на 25% лучшее, чем алгоритм JB2 при всех разрешениях сканирования в диапазоне 200 – 600 dpi.

**Выводы.** В настоящее время самыми мощными алгоритмами сжатия двухцветного изображения текста являются те, которые используют классификацию выделенных символов. При этом конечный результат – степень сжатия изображения – больше всего зависит от качества классификации, то есть количества полученных классов при неизменном условии, что в каждый класс входят изображения только одного символа. Еще один аспект при обсуждении любого метода сжатия – качество восстановленного изображения по сравнению с исходным сжимаемым оригиналом. Обычно, чем выше степень сжатия, тем

это качество хуже. Методы сжатия сканированных изображений текста, основанные на классификации выделенных символов, позволяют получать восстановленные изображения символов более высокого качества, чем оригинальные. Причем, чем лучше проведена классификация, тем больше сжатие и тем лучше качество изображения символов. Этот парадоксальный факт объясняется очень просто. Подходящая статистическая обработка этого класса позволяет избавиться от искажений, внесенных при печати и сканировании и имеющих случайный характер.

Описанные выше меры отличия, учитывающие контурный характер шумов печати и сканирования, являются достаточно сложными. Но все же есть уверенность, что они будут существенно улучшены, так чтобы проводимая с их помощью классификация выделенных изображений была близка к идеальной.



## Перелік використаної літератури

1. Азаров С. С. Современные модели провайдинга / С. С. Азаров, В. А. Хорошко. – К. : Изд-во «Полиграф Консалтинг», 2006. – 98 с.
2. Азаров С. С. Типизация моделей технологических решений провайдинга Интернет / С. С. Азаров, Е. Б. Азаров, М. М. Дивизинюк, В. А. Хорошко. – Севастополь : СНУЯЭиП, 2005. – 112 с.
3. Айвазян С. А. Прикладная статистика: Классификация и снижение размерности [Текст] / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков и др. – М. : Финансы и статистика, 1989. – 607 с.
4. Аксак Н. Г. Мультиагентная модель адаптации медицинского Интернет-ресурса на основе информационного портрета пользователя / Н. Г. Аксак, С. А. Коргут, И. В. Новосельцев // Системи обробки інформації. – Х. : ХУПС, 2013. – Вип. 2 (109). – С. 242–246.
5. Алексієв В. О. Управління розвитком транспортних систем (автоматика, телематика та мехатроніка на автомобільному транспорті) ВНЗ / В. О. Алексієв. – Харків : ХНАДУ, 2008. – 268 с.
6. Балдин К. В. Управленческие решения: учебник / К. В. Балдин, С. Н. Воробьев, В. Б. Уткин. – 4-е изд. – М. : Издательско-торговая корпорация «Дашков и К», 2014. – 496 с.
7. Балшикеева К. С. Современное геодезическое обеспечение внедрения трехмерного лазерного сканирования в строительстве [Текст] / К. С. Балшикеева // Молодой ученый. – 2015. – №11. – С. 264–266.
8. Будников С. А. Полумарковская модель сложного конфликта радиоэлектронных систем [Текст] / С. А. Будников // V Межд. конф. «Методы и средства управления технологическими процессами», Саранск, 19–21 ноября 2009 года. – Режим доступа : <http://fetmag.mrsu.ru/2009-2/>.
9. Букашкин С. А. Применение цифровых процессоров обработки сигналов – магистральный путь развития современных средств телекоммуникаций / С. А. Букашкин, Г. Ф. Лисицын, В. Г. Миронов // Труды 3-ей Международной конференции "Цифровая обработка сигналов и ее применение". – М. : Инсвязьиздат, 2000. – 315 с.
10. Вильхивская О. В. Методика планирования внедрения технологий электронного бизнеса на предприятиях машиностроительной отрасли / О. В. Вильхивская, Н. А. Брынза // Научные исследования: от теории к

практике: материалы III Междунар. науч. - практ. конф. (30.04.2015 г.). – Чебоксары : ЦНС «Интерактив плюс», 2015. – Т. 2. – С. 207–215.

11. Вихров А. А. Интеграция систем KPI/BSC и бюджетирования в единый инструмент управления / А. А. Вихров, П. А. Лекомцев // Управление компанией. – 2005. – № 10 (53). – С. 15. – Режим доступа : [http://iteam.ru/publications/finances/section\\_12/article\\_2634](http://iteam.ru/publications/finances/section_12/article_2634). – Дата обращения: 15 ноября 2016.

12. Вільхівська О. В. Методичний підхід до вибору технологій електронного бізнесу на підприємстві / О. В. Вільхівська // Системи обробки інформації : збірник наукових праць. – Х. : Харківський університет Повітряних Сил імені Івана Кожедуба, 2016. – Вип. 5 (142). – С. 228.

13. Вітлінський В. В. Моделювання економіки : навч. посіб. / В. В. Вітлінський. – К. : КНЕУ, 2003. – 408 с.

14. Галицкий А. В. Защита информации в сети – анализ технологий и синтез решений / А. В. Галицкий, С. Д. Рябко, В. Ф. Шаньгин. – М. : ДМК Пресс, 2004. – 616 с.

15. Герасименко В. А. Защита информации в автоматизированных системах обработки данных. В 2-х кн. Кн. 1 / В. А. Герасименко. – М., 1994. – 400 с.

16. Гнеденко Б. В. Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. – Изд. 3-е, испр. и доп. – М. : Комкнига, 2005. – 400 с.

17. Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.

18. Грушо А. А. Теоретические основы защиты информации / А. А. Грушо, Е. Е. Тимонина. – М. : Яхтсмен, 1996. – 192 с.

19. Деклараційний патент України на винахід № 56943А, Україна, МПК H03G5/08. Фільтр нижніх частот для підвищення стійкості і точності кутового автосупроводження літальних апаратів / О. В. Коломійцев, Г. В. Альошин, Д. П. Пашков – № 2002129790; заяв. 06.12.2002; опубл. 15.05.2003; Бюл. № 5. – 4 с.

20. Доповідь про стан інформатизації та розвиток інформаційного суспільства в Україні за 2014 рік [Електронний ресурс] // Державне агентство з питань науки, інновацій та інформатизації в Україні. – Режим доступу : <http://www.dkni.gov.ua/content/shchorichna-dopovid-pro-rozvytok-informaciyного-suspilstva>.

21. Дубов Д. В. Інформаційне суспільство в Україні: глобальні виклики та національні можливості : аналітична доповідь / Д. В. Дубов, М. А. Ожеван, С. Л. Гнатюк. – К. : НІСД, 2010. – 29 с.
22. Дюбуа Д. Теория возможностей. Приложения к представлению знаний в информатике / Д. Дюбуа, А. Прад // Перевод с фр. – М. : Радио и связь, 1990. – 288 с.
23. Евсеев С. П. Методология оценивания безопасности информационных технологий автоматизированных банковских систем Украины / С. П. Евсеев // *Ukrainian Scientific Journal of Information Security*, 2016. – vol. 22, issue 3. – pp. 297–309.
24. Єжова Л. Ф. Управління інформаційною безпекою. В 2-х томах / Л. Ф. Єжова, І. О. Мачалін, Я. В. Невойт, В. О. Хорошко. – К. : Вид. ДУІКТ, 2010. – 201 с.
25. Жуков В. Г. Модель нарушителя прав доступа в автоматизированной системе / В. Г. Жуков, М. Н. Жукова, А. П. Стефаров // Программные продукты и системы. – 2012. – № 2. – С. 75 – 78.
26. Засядько А. А. Дослідження ефективності роботи виробничого підприємства / А. А. Засядько, С. С. Королюк // *Матеріали International Scientific-Practical Conference Economic Development Strategy in Terms of European Integration.*, Kaunas, 2016. – 341 p.
27. Затхей В. А. Формалізація знань в системі дистанційного навчання / В. А. Затхей, І. П. Ковріжних // Системи обробки інформації. – Х. : Харківський університет Повітряних Сил імені Івана Кожедуба, 2012. – Вип. 8(106) – С. 242.
28. Затхей В. А. Использование функциональных сетей для означивания начальных условий в открытых экспертных системах / В. А. Затхей, С. Є. Селезньов // *Збірник наукових праць / ХВУ.* – Х., 2000. – Вип. 4(30). – С. 126–131.
29. Затхей В. А. Формально-логический аппарат представления знаний о процессах управления обучением в экспертных обучающих системах / В. А. Затхей, Н. В. Шаронова, И. Е. Лещенко // «АСУ и приборы автоматизации» № 130 / М-во образования и науки Украины, Харьковский национальный университет радиоэлектроники. – Х., 2005. – С. 52–56.
30. Зегджа Д. П. Основы безопасности информационных систем / Д. П. Зегджа, А. М. Ивашко. – М. : Горячая линия – Телеком, 2000. – 452 с.

31. Иванов В. Г. Сжатие изображения текста на основе статистического анализа и классификации вертикальных элементов строки [Текст] / В. Г. Иванов, Ю. В. Ломоносов, М. Г. Любарский // Восточно-Европейский журнал передовых технологий. – Харьков, 2014. – № 4/2 (70). – С. 4–15.
32. Иванов В. Г. Сжатие изображения текста на основе выделения символов и их классификации [Текст] / В. Г. Иванов, М. Г. Любарский, Ю. В. Ломоносов // Проблемы управления и информатики. – 2010. – № 6. – С. 111–122.
33. Иванов В. Г. Сжатие изображения текста на основе формирования и классификации вертикальных элементов строки в графическом словаре символьных данных [Текст] / В. Г. Иванов, М. Г. Любарский, Ю. В. Ломоносов // Проблемы управления и информатики. – 2011. – № 5. – С. 98–109.
34. Иванов В. Г. Формальное описание дискретных преобразований Хаара // Проблемы управления и информатики. – 2003. – № 5. – С. 68–75.
35. Иванов В. Г. Сокращение содержательной избыточности изображений на основе классификации объектов и фона / В. Г. Иванов, М. Г. Любарский, Ю. В. Ломоносов // Проблемы управления и информатики. – Київ, 2007. – № 3. – С. 93–102.
36. Информационные технологии и защита информации в информационно-коммуникационных системах : коллективная монография / С. П. Евсеев, М. Ю. Лосев, С. В. Минухин и др. ; под ред. В. С. Пономаренко. – Х. : Вид-во ТОВ «Щедра садиба плюс», 2015. – 486 с.
37. Информационные технологии и системы в управлении, образовании, науке : коллективная монография / А. В. Коломийцев и др. ; под ред. В. С. Пономаренко. – Х. : Цифрова друкарня № 1, 2013. – 278 с.
38. Информационные технологии и системы в управлении, образовании, науке : коллективная монография / С. П. Евсеев, М. Ю. Лосев, С. В. Минухин и др. ; под ред. В. С. Пономаренко. – Х. : Вид-во ТОВ «Цифрова друкарня № 1». – 2013. – 278 с.
39. Камер Дуглас Э. Сети TCP/IP, том 1. Принципы, протоколы и структура / Камер Дуглас Э. – М.: Издательский дом "Вильямс", 2003. – 445с.
40. Каплан Р. Стратегические карты. Трансформация нематериальных активов в материальные результаты / Р. Каплан, Д. П. Нортон. – М. : ЗАО «Олимп-Бизнес», 2005. – 512 с.

41. Кейнс Дж. М. Общая теория занятости, процента и денег / Дж. М. Кейнс. – М. : Гелиос АРВ, 2002. – 352 с.
42. Коваленко А. А. Подходы к оптимизации распределения задач управления по компонентам компьютерной системы, образующей систему управления объектом критического применения / А. А. Коваленко // Наука у техніка Повітряних Сил Збройних Сил України: науково-технічний журнал. – Х. : ХУ ПС, 2014. – Вип. 2(15). – С. 158–160.
43. Коваленко А. А. Подходы к синтезу информационной структуры системы управления объектом критического применения / А. А. Коваленко // Системы обработки информации : сборник научных трудов. – Х. : ХУ ВС, 2014. – Вип. 1 (117). – С. 180–184.
44. Коваленко А. А. Подходы к синтезу технической структуры компьютерной системы, образующей систему управления объектом критического применения / А. А. Коваленко // Сборник научных трудов Харьковского университета Воздушных Сил. – Х. : ХУ ВС, 2014. – Вип. 1 (38). – С. 116–119.
45. Коваленко А. В. Метод количественной оценки рисков разработки программного обеспечения / А. А. Смирнов, А. В. Коваленко, Н. Н. Якименко, А. П. Доренский // Збірник наукових праць Харківського університету Повітряних Сил. – Харків : ХУПС, 2016. – Вип. 2 (47). – С. 128–133.
46. Коломійцев О. В. Адаптована структура приймально-передавальної частки вимірювальних каналів для синтезу лазерної інформаційно-вимірювальної системи з використанням частотно-часового методу / О. В. Коломійцев // Системи обробки інформації. – Х. : ХУ ПС, 2011. – Вип. 4 (94). – С. 28–31.
47. Компьютерное зрение. Современный подход / Д. Форсайт, Д. Понс ; пер. с англ. – М. : Вильямс, 2004. – 928 с.
48. Корнеев В. Будущее высокопроизводительных вычислительных систем // Открытые системы. – М., 2003. – № 5. – С. 10–17.
49. Косяков А. Системная инженерия. Принципы и практика / под ред. В. К. Батоврина. – М. : ДМК Пресс, 2014. – 624 с.
50. Кунт М. Методы кодирования изображений второго поколения / М. Кунт, А. Икономопулос, М. Кошер // ТИИЭР. – 1985. – Т. 73, № 4. – С. 59–86.
51. Кучук Г. А. Концептуальний підхід до синтезу структури інформаційно-телекомунікаційної мережі / Г. А. Кучук, І. В. Рубан,

О. П. Давікоза // Системи обробки інформації : збірник наукових праць. – Х. : ХУ ПС, 2013. – Вип. 7 (114). – С. 106–112.

52. Кучук Г. А. Синтез стратифікованої інформаційної структури інтеграційної компоненти гетеро-генної складової Єдиної АСУ Збройними Силами України / Г. А. Кучук, О. П. Давікоза // Наука і техніка Повітряних Сил Збройних Сил України: науково-технічний журнал. – Х. : ХУ ПС, 2013. – № 3 (12). – С. 154–158.

53. Ленков С. В. Методы и средства защиты информации. В 2-х томах / С. В. Ленков, Д. А. Перегудов, В. А. Хорошко. – К. : Арий, 2008. – 344 с.

54. Листровой С. В. Модель и подход к планированию распределения ресурсов в гетерогенных Грид-системах / С. В. Листровой, С. В. Минухин // Международный научно-технический журнал "Проблемы управления и информатики". – 2012. – № 5. – С. 120–133.

55. Литвиненко К. В. Полумарковский гиперслучайный подход к оценке рисков систем // Збірник наукових праць. – ОДАТРЯ, 2014. – № 1 (4). – С. 77–80.

56. Лосев Ю. И. Автоматизация в сетях с коммутацией пакетов / Ю. И. Лосев, М. Ю. Лосев, Ф. К. Яковец. – К. : «Техніка», 1994. – 212 с.

57. Маракова И. И. Синтез и исследование методов верификации объектов электронного документооборота / И. И. Маракова, Л. А. Кузнецова, А. А. Сыропятов // Захист інформації. – 2008. – № 2. – С. 50–65.

58. Маракова І. І. Технологія цифрових водяних меток з головними покриваючими повідомленнями в нагляді бінарних зображень // Правове, нормативне та метрологічне забезпечення систем захисту інформації в Україні. – Науково-технічний збірник. – К. : НДЦ „Тезис” НТУУ „КПІ”, 2003. – Вип. 7. – С. 53–58.

59. Милославская Н. Г. Интрасети: доступ в Internet, защита : учебное пособие для вузов / Н. Г. Милославская, А. И.-Толстой. – М. : ЮНИТИ – ДАНА, 2000. – 527 с.

60. Минухин С. В. Имитационная модель и ее программная реализация планирования ресурсов Грид-системы / С. В. Минухин, С. В. Знахур // Проблеми програмування. – 2012. – № 2-3. Спеціальний випуск. – С. 133–142.

61. Минухин С. В. Исследование и анализ основанных на методе покрытия алгоритмов планирования для гетерогенной распределенной системы с использованием сценарного подхода / С. В. Минухин, С. В. Знахур //

Информационные технологии и системы в управлении, образовании, науке : монография. – Х. : Цифрова друкарня, 2013. – С. 39–54.

62. Минухин С. В. Модели и методы решения задач планирования в распределенных вычислительных системах : монография / С. В. Минухин. – Х. : Изд-во ООО "Щедрая усадьба плюс", 2014. – 324 с.

63. Міхайленко В. М. Математичний аналіз для економістів. / В. М. Міхайленко, Н. Д. Федоренко–К. : Вид-во ЄУФІМБ 2002. – 297 с.

64. Моделі визначення компетентностей у системі дистанційного навчання : монографія / В. П. Степанов, І. О. Борозенець, В. П. Бурдаєв та ін. ; за заг. ред. В. П. Степанова. – Х. : Вид. ХНЕУ, 2013. – 224 с.

65. Мокрышев А. Основные функции ETL-систем / А. Мокрышев // Хабрахабр: Разработка, 19 января 2015. – Режим доступа : <https://habrahabr.ru/post/248231>. – Дата обращения: 18 ноября 2016.

66. Москаленко В. В. Структура системы Enterprise Performance Management с учётом технологии каскадирования ключевых показателей деятельности / В. В. Москаленко, Н. Г. Фонта // Вісник Національного технічного університету «ХПІ». (Серія «Системний аналіз, управління та інформаційні технології»). – Харків: НТУ «ХПІ», 2016. – № 45. – С. 34–40.

67. Ніконов О. Я. Розроблення та впровадження інтернет-технологій для підвищення ефективності використання транспортних засобів / О. Я. Ніконов, В. О. Алексієв, В. Ю. Улько, Г. І. Середіна // Вісник СевНТУ. – 2013. – Вип. 142. – С. 69–72.

68. Новосельцев И. В. Показатели эффективности нейросетевой системы медицинской диагностики / И. В. Новосельцев, Н. Г. Аксак // Системи управління, навігації та зв'язку. – Київ, 2010. – Вип. 4 (16). – С. 142–149.

69. Новосельцев И. В. Система первичной диагностики меланомы кожи / И. В. Новосельцев, Н. Г. Аксак // Бионика интеллекта : научн.-техн. журнал. – 2010. – № 3 (74). – С. 94–98.

70. Нурдинов Р. А. Подходы и методы обоснования целесообразности выбора средств защиты информации [Электронный ресурс] / Р. А. Нурдинов, Т. Н. Батова // Современные проблемы науки и образования. – 2013. – № 2. – Режим доступа к ресурсу : <http://elibrary.ru/item.asp?id=21285749>.

71. Палагин А. В. Реконфигурируемые вычислительные системы: Основы и приложения / А. В. Палагин, В. Н. Опанасенко. – К. : Просвіта, 2006. – 280 с.

72. Патент на корисну модель № 43725, Україна, МПК H04 Q 1/453. Модифікований селектор подовжніх мод / О. В. Коломійцев, Г. В. Альошин, В. В. Белімов та ін. – № u200903693; заяв. 15.04.2009; опубл. 25.08.2009; Бюл. № 16. – 6 с.

73. Петров А. А. Компьютерная безопасность. Криптографические методы защиты информации / А. А. Петров. – М. : ДМК, 2000. – 448 с.

74. Про основні засади розвитку інформаційного суспільства в Україні на 2007-2015 роки: Закон України №537-V від 9 січня 2007 року // Відомості Верховної Ради України (ВВР). – 2007. – № 12. – ст. 102

75. Про схвалення Стратегії розвитку інформаційного суспільства в Україні: Розпорядження Кабінету Міністрів №386-р від 15 травня 2013 року. [Електронний ресурс]. – Режим доступу : <http://zakon2.rada.gov.ua/laws/show/386-2013-p>.

76. Проект Закону України “Про основні засади забезпечення кібербезпеки України” від 19.06.2015 № 2126а. – [Електронний ресурс]. – Режим доступу : [http://w1.c1.rada.gov.ua/pls/zweb2/webproc4\\_1?pf3511=55657](http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=55657).

77. Проект Стратегії забезпечення кібернетичної безпеки України. – [Електронний ресурс]. – Режим доступу : [http://www.niss.gov.ua/public/File/2013\\_nauk\\_an\\_rozrobku/kiberstrateg.pdf](http://www.niss.gov.ua/public/File/2013_nauk_an_rozrobku/kiberstrateg.pdf).

78. Пушкар О. І. Оцінка вартості розробки та впровадження технологій електронного бізнесу на підприємстві / О. І. Пушкар, О. В. Вільхівська // Бізнес-інформ. – Харків, 2016. – № 12. – С. 38–44.

79. Риз Дж. Облачные вычисления : пер. с англ. – СПб. : БХВ-Петербург, 2011. – 288 с.

80. Ричардсон Ян. Видеокодирование. H. 264 и MPEG-4 – стандарты нового поколения : пер. с англ. – М. : Техносфера, 2005. – 368 с.

81. Різник В. В. Синтез оптимальних комбінаторних систем. – Львів : Вища школа, 1989. – 165 с.

82. Різник О. Я. Метод побудови завадостійких кодів на основі ідеальних кільцевих відношень / О. Я. Різник, О. В. Повшук, Ю. Є. Кинаш, Б. І. Балич // Тези Міжнародної наукової конференції “Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту” (ISDMCI'2015), 25–28 травня 2015 р., Залізний Порт. – С. 209–211.

83. Різник О. Я. Авторський захист графічних зображень за допомогою числових в'язанок / О. Я. Різник, Б. І. Балич, Д. Ю. Скрибайло-Леськів // Вісник



НУ "Львівська політехніка" "Комп'ютерні науки та інформаційні технології". – 2008. – С. 62–64.

84. Різник О. Я. Дослідження фільтрів для обробки зображень з використанням ідеальних кільцевих в'язанок / О. Я. Різник, А. Єрьоменко // Вісник НУ"ЛП" Комп'ютерні науки та інформаційні технології. – Львів : Видавництво НУ "Львівська політехніка", 2011. – № 719. – С. 13–19.

85. Різник О. Я. Інформаційні технології оптимізації структур технічних систем на основі нееквідистантних конфігурацій / О. Я. Різник, Ю. Є. Кинаш, Н. О. Кустра // Тези Міжнародної наукової конференції Присвяченої пам'яті професора Рогальського Ф.Б. "Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту" (ISDMCI'2014), 28–31 травня 2014 р. – Залізний Порт, 2014. – С. 157–160.

86. РС БС ИББС – 2.2-2009. Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Методика оценки рисков нарушения информационной безопасности – [Электронный ресурс]. – Режим доступа к ресурсу : [www.cbr.ru/credit/gubzi\\_docs/st22\\_09.pdf](http://www.cbr.ru/credit/gubzi_docs/st22_09.pdf).

87. Руденко О. Г. Козволюционирующие нейронные сети прямого распространения / О. Г. Руденко, А. А. Бессонов // Проблемы управления и информатики. – 2016. – № 5. – С. 63–73.

88. Сверхбольшие интегральные схемы и современная обработка сигналов: пер. с англ. / под ред. : С. Гуна, Х. Уайтхадса, Т. Кайлата. – М. : Радио и связь, 1989. – 472 с.

89. Теорія оптимальних рішень. Моделювання та керування в умовах невизначеності : зб. наук. пр. / ред. : Т. П. Мар'янович ; НАН України. Ін-т кібернетики ім. В. М. Глушкова. – К., 2000. – 174 с.

90. Федоренко Ю. С. Кластеризация данных на основе нейронного газа и марковских алгоритмов / Ю. С. Федоренко // Молодежный научно-технический вестник МГТУ им. Н. Э. Баумана. Электрон. журн. – 2014. – № 8. – Режим доступа : <http://sntbul.bmstu.ru/doc/730611.html>.

91. Филимонов А. Ю. Протоколы Интернета. – СПб. : БХВ-Петербург, 2008. – 528с.

92. Хома І. Б. Економіко-математичні методи аналізу діяльності підприємств / І. Б. Хома, В. В. Турко ; Нац. ун-т « Львів. Політехніка». – Л. : Вид-во Нац. ун-ту «Львів. Політехніка», 2008. – 328 с.

93. Хоффман Л. Современные методы защиты информации / Л. Хоффман. – М. : Сов. радио, 1980. – 264 с.
94. Чалая Л. Э. Оценивание пертинентности лингвистических дескрипторов в системах информационного поиска документов [Текст] / Л. Э. Чалая, Ю. Ю. Харитонова // Восточно-европейский журнал передовых технологий. – 2015. – № 1/9 (73). – С. 46–53.
95. Чалая Л. Э. Метод двухэтапной классификации электронных текстов [Текст] // Л. Э. Чалая, С. Г. Удовенко, Е. С. Кушвид // Бионіка інтелекту. – 2016. – № 2 (87). – С. 16–23.
96. Шинкарук Л. В. Структурні трансформації в економіці України: динаміка, суперечності та вплив на економічний розвиток : наукова доповідь / Л. В. Шинкарук, І. А. Бевз, І. В. Барановська та ін. ; за ред. чл.-кор. НАН України Л. В. Шинкарук ; НАН України, ДУ «Ін-т екон. та прогноз. НАН України». – К., 2015. – 304 с.
97. Шлезингер М. И. Математические средства обработки изображений [Текст] / М. И. Шлезингер. – Киев : Наукова думка, 1983. – 200 с.
98. Юдін О. К. Методологія побудови класифікатора загроз державним інформаційним ресурсам / О. К. Юдін, С. С. Бучик, А. В. Чунарьова, О. І. Варченко // Наукоємні технології. – 2014. – № 2 (22). – С. 200-210.
99. Ярочкин В. И. Безопасность информационных систем / В. И. Ярочкин. – М. : Ось-89, 1996. – 320 с.
100. Ярочкин В. И. Служба безопасности коммерческого предприятия / В. И. Ярочкин. – М. : Ось-89, 1995. – 144 с.
101. 3D редакторы, плюсы и минусы / Хабрахабр [Электронный ресурс]. – Режим доступа : <http://habrahabr.ru/post/136350/>.
102. Axak N. G. Development of multi-agent system of neural network diagnostics and remote monitoring of patient / N. G. Axak // Eastern-European Journal of Enterprise Technologies. – 2016. – 4/9 ( 82 ) – P. 4-11.
103. Blobel B. (2006). Advanced and secure architectural \* Encyclopedia of Healthcare Information Systems MEDICAL INFORMATION SCIENCE REFERENCE. – New York, 2008. – 231p.
104. Brian Evans. Practical 3D Printers: The Science and Art of 3D Printing / Evans Brian. – New York : Apress, 2012. – 332 p.

105. Coello C. A. Multi-objective optimization of trusses using genetic algorithms / C. A. Coello, A. D. Christiansen // *Computers & Structures*. – 2000. – 75. – P. 647–660.
106. Donkor E. Urban Water Demand Forecasting: Review of Methods and Models / Donkor E., Mazzuchi T., Soyer R., Roberson A. // *Journal of Water Resources Planning and Management*. – 2014. – № 140 (2). – P. 146–159.
107. Fibich P. Model of grid scheduling problem / P. Fibich, L. Matyska, H. Rudová // *Exploring Planning and Scheduling for Web Services, Grid and Autonomic Computing*, AAAI Press, 2005. – P. 17–24.
108. Furoo S. An enhanced self-organizing incremental neural network for online unsupervised learning / S. Furoo, T. Ogura, O. Hasegawa // *Neural Networks*. – 2007. – № 6. – pp. 893-903.
109. Furoo S. An incremental network for on-line unsupervised classification and topology learning / S. Furoo, O. Hasegawa // *Neural Networks*. – 2006. – Vol. 19, № 1. – pp. 90-106.
110. Garcia-Pedrajas N. Covnet: A cooperative coevolutionary model for evolving artificial neural networks / N. Garcia-Pedrajas, C. Hervas-Martinez, J. Munoz-Perez // *IEEE Transactions on Neural Networks*. – 2003. – 14 (3). – P. 575–596.
111. Holland J. *Adaptation in Natural and Artificial Systems* 2nd edn. – Cambridge, MIT Press. – 1992. – 228 P.
112. Horielova K. A. Long-term and medium-term forecasting of water consumption of large cities / K. A. Horielova, V. M. Zadachyn // *Збірник наукових праць “Системи обробки інформації”*. – Харків : Вид. ХУПС, 2016. – Вип. № 1 (138). – С. 76-80.
113. Hryshchuk R. The synergetic approach for providing bank information security: the problem formulation // R. Hryshchuk, S. Yevseiev / *Безпека інформації*. – 2016. – № 22 (1). – С. 64–74. – doi:10.18372/2225-5036.22.10456.
114. Idris F. Review of Image and Video Indexing Techniques / F. Idris, S. Panchanathan // *Journal of Visual Communication and Image Representation*, 1997. – Vol. 8. – pp. 53–73.
115. Junguo L. Forecast of Water Demand in Weinan City in China using WDF-ANN model / L. Junguo, H. Savenije, J. Xu // *Physics and Chemistry of the Earth*. – 2003. – № 28. – P. 219–224.
116. Kudriashov V. 'Experimental Evaluation of Opportunity to Improve the Resolution of the Acoustic Maps'. In: Kountchev R. and Nakamatsu K. (eds.), *New*

Approaches in Intelligent Image Analysis, Intelligent Systems Reference Library 108, pp. 353-373. Springer International Publishing Switzerland 2016. DOI: 10.1007/978-3-319-32192-9\_11, SJR: 0.154.

117. Moulin P., O'Sullivan. Information-theoretic Analysis of Watermarking // Proc. of the International Conference on Acoustic, Speech and Signal Processing. – 2000. – Vol. 6. – P. 3630–3633.

118. MySQL. Оптимизация производительности, 2-е издание / Шварц Б., Зайцев П., Ткаченко В., Заводны Дж., Ленц А., Беллинг Д. ; пер. з англ. – СПб. : Символ-Плюс, 2010. – 832 с.

119. Semi-Markov risk models for finance, insurance and reliability [Electronic resource] / J. Jacques, M. Raimondo. - Electronic text data. - Boston, Ma : Springer Science + Business Media LLC, 2007.

120. Shklovets A. V. Visualization of High Dimensional Data Using Two Dimensional Self Organizing Piecewise Smooth Kohonen Maps /A. V. Shklovets and N. G. Ахак //ISSN 1060 992X, Optical Memory and Neural Networks (Information Optics), 2012. – Vol. 21, No. 4. – pp. 227–232.

121. Technical Papers from AT&T Labs: Электронный ресурс. – Режим доступа : <http://dvvuzone.org/techpapers/index.html>.

122. The Global Competitiveness Report 2012–2016 // [Электронный ресурс]. – Режим доступа : <https://www.weforum.org/>.

123. The Global Information Technology Report 2012 -2016 // [Электронный ресурс]. – Режим доступа : <https://www.weforum.org/>.

124. The Ultimate Guide to 3D Printing London: MagBook, 2014. – 148 p.

125. Tiwari M. K. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models / M. K. Tiwari, J. Adamowski // Water Resources Research. – 2013. – vol. 49. – P. 6486–6507.

126. Top 25: Most Popular 3D Modeling Software for 3D Printing [Электронный ресурс] – Режим доступа: <https://i.materialise.com/blog/top-25-most-popular-3d-modeling-design-software-for-3d-printing>.

127. Yun Bai A multi-scale relevance vector regression approach for daily urban water demand forecasting / Yun Bai, Pu Wang, Chuan Li, Jingjing Xie, Yin Wang // Journal of Hydrology. – 2014. – vol. 517. – P. 236–245.

128. Zitzler E. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach / E. Zitzler, L. Thiele // IEEE Trans. on Evolutionary Computation. – 1999. – 3, N 4. – P. 257–271.

129. Petryshyn L. B. Theory of digital data processing in the ICT. In monography *Advances in ICT for Business, Industry and Public Sector*. Springer International Publishing Switzerland. 2015. –pp. 157-170.

130. Kehtarnavaz N. *Real-Time Digital Signal Processing*. Elsevier Inc. 2011. – 320 p.

131. Morgera S. *Digital Signal Processing: Applications to Communications and Algebraic Coding Theories*. Elsevier Inc. 2012. – 254 p.

132. Petryshyn L. Pelech-Pilichowski T. On a Property of Phase Correlation and Possibilities to Reduce the Walsh Function System // In monography: *Advances in Business ICT: New Ideas from Ongoing Research*. / ed-s Tomasz Pelech-Pilichowski, Maria Mach-Krol, Celina M. Olszak. *Studies in Computational Intelligence 658*. Springer International Publishing AG 2017, Cham, Switzerland. — P. 125–135.

133. Petryshyn L. Applying of the Walsh Functions Systems in Navigation Digital Data Processing. 4th International Conference *Methods and Systems of Navigation and Motion Control (MSNMC-2016)*. October 18-20, 2016. – P. 300-304.

134. Стахов А. П. Введение в алгоритмическую теорию измерения / А. П. Стахов. – М. : Советское Радио, 1977. – 288 с.

135. Петришин М. Л. Застосування векторно-розгалужуючих схем в моделюванні процесів ПФІ // II Міжнародна конференція “Комп’ютерна алгебра та інформаційні технології” : тези доповідей, 21-26 серпня 2016. / Одеський національний університет імені І. І. Мечникова. – Одеса, 2016. – 60с.

136. Брянский Л. Н. *Метрология. Шкалы, эталоны, практика*. – М. : ВНИИФТРИ, 2004. – 222 с.

137. Петришин Л. Б. Моделювання субтрактивно-адитивного способу перетворення форми інформації / Л. Б Петришин // *Математичний вісник НТШ* ; ISSN 1812 - 6774. – 2012. – т. 9 – P. 246–268. — Bibliogr. – P. 266–268.

**Міністерство освіти і науки України**

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ:  
ПРОБЛЕМИ ТА ПЕРСПЕКТИВИ**

*Монографія*

*За загальною редакцією*

*д-ра екон. наук, професора В. С. Пономаренка*

**Харків, 2017**

УДК 681.518.54

ББК 32.966

И 74

*Рекомендовано до видання рішенням ученої ради Харківського національного економічного університету імені Семена Кузнеця  
(протокол № 6 від 06.03.2017 р.)*

**Рецензенти:**

**Корабльов Микола Михайлович** – доктор технічних наук, професор, професор кафедри ЕОМ Харківського національного університету радіоелектроніки (ХНУРЕ);

**Семенов Сергій Генадійович** – доктор технічних наук, завідувач кафедри “Обчислювальної техніки і програмування” НТУ “ХПІ”;

**Остапов Сергій Едуардович** – доктор фізико-математичних наук, завідувач кафедри програмного забезпечення комп’ютерних систем Чернівецького національного університету ім. Ю. Федьковича.

**И 74** Інформаційні технології: проблеми та перспективи : монографія / за заг. ред. В. С. Пономаренка. – Х. : Вид. Рожко С. Г., 2017. – 447 с.  
ISBN 978-966-97498-5-7

Розглянуто й обґрунтовано результати наукових досліджень в галузі розробки і практичного застосування сучасних інформаційних технологій.

Монографія представляє інтерес як для фахівців, сфера діяльності яких безпосередньо пов’язана з розробкою ІТ-технологій, способів забезпечення безпеки і передачі в комунікаційних системах, так і для більш широкого кола фахівців. Вона буде корисною викладачам, аспірантам і студентам, що спеціалізуються в області ІТ-технологій, і всім, хто серйозно цікавиться проблемами взаємодії інформаційних технологій і суспільства.

Рассмотрены и обоснованы результаты научных исследований в области разработки и практического применения современных информационных технологий.

Монография представляет интерес как для специалистов, сфера деятельности которых непосредственно связана с разработкой ИТ-технологий, способов обеспечения безопасности и передачи в коммуникационных системах, так и для более широкого круга специалистов. Она будет полезна преподавателям, аспирантам и студентам, специализирующимся в области ИТ-технологий, и всем, кто серьезно интересуется проблемами взаимодействия информационных технологий и общества.

*За достовірність викладених фактів, цитат та інших відомостей  
відповідальність несе автор.*

ISBN 978-966-97498-5-7

© Коллектив авторов, 2017

## ЗМІСТ

Вступ	5
<i>Інформаційні технології в технічних системах</i>	
Розділ 1. Альошин Г. В., Коломійцев О. В. Проблеми створення багатофункціональної лазерної системи контролю і управління літальним апаратом	16
Розділ 2. Алексієв В. О., Алексієв О. П. Інформаційний розвиток порталу віртуального управління процесами транспортного обслуговування	32
Розділ 3. Іванов В. Г. Параллельные и последовательные структуры Хаара для цифровой обработки сигналов	48
Розділ 4. Коваленко А. А. Синтез інформаційної та технічної структури системи управління об'єктом критичного застосування	64
Розділ 5. Іванов В. Г. Сжатие данных портретных изображений	78
Розділ 6. Руденко О. Г., Безсонов О. О. Многокритериальная нейроидентификация нелинейных объектов на основе коэволюционного подхода	94
Розділ 7. Мінухін С. В. Модель планирования пакетов заданий в распределенной вычислительной системе	110
Розділ 8. Лосев М. Ю. Задачи и алгоритмы управления маршрутизацией пакетов в сетях	120
Розділ 9. Різник О. Я. Фільтри графічних зображень на основі числових в'язанок	137
Розділ 10. Удовенко С. Г., Чала Л. Е. Нейросетевой метод кластеризации электронных текстов в распределенной файловой системе	154
<i>Захист інформації в інформаційних та комунікаційних системах</i>	
Розділ 11. Дудикевич В. Б., Микитин Г. В. Комплексні системи безпеки кібернетичного простору кіберфізичної системи на основі концепції “об'єкт – загроза – захист”	170
Розділ 12. Хорошко В. О., Хохлачова Ю. Є. Підхід до вибору засобів захисту з урахуванням їх оцінок	187
Розділ 13. Євсєєв С. П., Король О. Г. Синергетические модели оценки безопасности в автоматизированных банковских системах	203



Розділ 14. Мельник М. О., Нікітін Г. Д. Моделі побудови політики інформаційної безпеки підприємства	222
Розділ 15. Маракова І. І., Кузнецова Л. А. Разработка и исследование методов реверсивных цифровых знаков (ЦВЗ) для верификации медицинских изображений	237
Розділ 16. Смірнов О. А., Коваленко О. В. Разработка метода управления рисками разработки программного обеспечения	254
Розділ 17. Петришин Л. Б. Аналіз системних властивостей та обґрунтування ефективності методів перетворення форми інформації	269
Розділ 18. Ровінський В. А., Петришин М. Л. Аналіз ефективності методів перетворення форми інформації в адитивних системах числення	281
<b><i>Інформаційні технології в економіці, екології, медицині та освіті</i></b>	
Розділ 19. Задачин В. М. Методи та моделі прогнозування міського добового водоспоживання	297
Розділ 20. Засядько А. А., Королюк С. С. Моделювання максимізації прибутку на основі виробничої функції Кобба-Дугласа	312
Розділ 21. Аксак Н. Г. Мультиагентная система нейросетевой диагностики и удаленного мониторинга пациента	325
Розділ 22. Коц Г. П., Моїсеєнко А. С. Розвиток ІТ-галузі в Україні: світові рейтинги, тенденції, перспективи	341
Розділ 23. Вільхівська О. В., Бринза Н. А. Определение стоимости разработки внедряемых технологий электронного бизнеса на промышленном предприятии	356
Розділ 24. Москаленко В. В., Фонта Н. Г. Подход к построению информационной системы управления эффективностью предприятия, входящего в холдинг	371
Розділ 25. Затхей В. А., Тесленко О. В. Формалізація знань про процеси підготовки вихідних даних при тестуванні	387
<b><i>Комп'ютеризовані технології та системи видавничо-поліграфічних виробництв та електронних мультимедійних видань</i></b>	
Розділ 26. Ізонін І. В., Лотошинська Н. Д. Дослідження технології побудови тривимірних моделей	402
Розділ 27. Ломоносов Ю. В. Методи підвищення якості класифікації символів при обробці текстових зображень	418
Перелік використаної літератури	434

## **Наукове видання**

*Аксак Наталія Георгіївна, Алексієв Олег Павлович, Алексієв Володимир Олександрович, Альошин Геннадій Васильович, Безсонов Олександр Олександрович, Бринза Наталя Олександрівна, Вільхівська Ольга Володимирівна, Дудикевич Валерій Богданович, Євсєєв Сергій Петрович, Задачин Віктор Михайлович, Засядько Аліна Анатоліївна, Затхей Володимир Анатолійович, Іванов Володимир Георгійович, Ізонін Іван Вікторович, Коваленко Андрій Анатолійович, Коваленко Олександр Володимирович, Коломійцев Олексій Володимирович, Королюк Сергій Степанович, Король Ольга Григорівна, Коц Григорій Павлович, Кузнецова Людмила Анатоліївна, Ломоносов Юрій Вячеславович, Лосєв Міхаїл Юрійович, Лотошинська Наталія Дмитрівна, Маракова Ірина Ільїнічна, Мельник Маргарита Олександрівна, Микитин Галина Василівна, Мінухін Сергій Володимирович, Моїсеєнко Аліна Сергіївна, Москаленко Валентина Володимирівна, Нікітін Геннадій Дмитрович, Петришин Любомир Богданович, Петришин Михайло Любомирович, Різник Олег Яремович, Ровінський Віктор Анатолійович, Руденко Олег Григорійович, Смірнов Олексій Анатолійович, Тєслєнко Олег Володимирович, Удовєнко Сергій Григорович, Фонта Наталія Григорівна, Хорошко Володимир Олексійович, Хохлачова Юлія Євгенівна, Чала Лариса Ернєстівна*

# **ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ: ПРОБЛЕМИ ТА ПЕРСПЕКТИВИ**

## **Монографія**

*За загальною редакцією д-ра екон. наук, професора В. С. Пономарєнка*

Підписано до друку 30.03.2017 р. Формат 60x84 1/16.  
Папір офсетний. Друк ризографічний. Ум. друк. арк. 26,0.  
Гарнітура Times New Roman. Наклад 300 прим. Зам. № 0330/8-17.

Видавець: Рожко Сергій Григорович  
Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру видавців,  
виготовлювачів і розповсюджувачів видавничої продукції

ДК № 4924 від 24.06.2015 р.

Адреса для листування: а/с 11437, м. Харків, 61171

Надруковано з готових оригінал-макетів у друкарні ФОП Петров В. В.  
Єдиний державний реєстр юридичних осіб та фізичних осіб-підприємців.

Запис № 2400000000106167 від 08.01.2009 р.

61144, м. Харків, вул. Гв. Широнінців, 79в, к. 137, тел. (057) 78-17-137.

e-mail:bookfabrik@mail.ua