

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

УДК 681.518.54



# Тези доповідей

**VII міжнародної науково-практичної  
конференції  
“Проблеми та перспективи розвитку  
ІТ-індустрії”  
17 – 18 квітня 2015 р.**

Харків 2015

**УДК 681.518.54**

Матеріали VII міжнародної науково-практичної конференції “Проблеми та перспективи розвитку ІТ-індустрії”: тези доповідей, 17 – 18 квітня 2015 р. – Х.: ХНЕУ імені Семена Кузнеця, 2015. – **85 с.**

Наведені тези пленарних та секційних доповідей за теоретичними та практичними результатами наукових досліджень і розробок. Представлені результати теоретичних досліджень в галузях проектування інформаційних систем, технологій захисту інформації, використання сучасних інформаційних технологій в управлінні системами, моделювання бізнес-процесів, застосування геоінформаційних технологій, дистанційній освіті, питань розвитку Start\_Up, інформаційних технологій в видавничо-поліграфічній галузі.

Матеріали публікуються в авторській редакції.

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

## АНАЛИЗ МЕТОДОВ СЖАТИЯ ИЗОБРАЖЕНИЙ ОЦИФРОВАННОГО ТЕКСТА

В работе рассматриваются методы классификации, применяемые при сжатии файла с битональным изображением текста, полученным сканированием или цифровым фотографированием.

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

Эффективность алгоритма в целом определяется *качеством классификации*, то есть количеством получившихся классов, которое в идеале должно совпадать с количеством символов, присутствующих в тексте, чье изображение сжимается [2].

Отмечено что главную трудность при классификации уже разделенных символов представляют шумы - случайные искажения, возникающие при печати страницы и ее последующем сканировании.

Особое внимание обращается на используемые при этом меры различия двух изображений символов, выделенных из изображения текста.

Чтобы разделить имеющуюся совокупность изображений символов на классы нужно:

- выбрать какую-либо *меру отличий* изображений двух символов, позволяющую, если она достаточно мала, утверждать, что изображения представляют собой один и тот же символ (soft matching), и их можно отнести к одному и тому же классу;

- выбрать какой-нибудь алгоритм разбиения на классы.

Мера отличия – это какая-либо функция, определенная для каждой пары элементов  $S_1$  и  $S_2$  классифицируемого множества. Эти меры различия позволяют с той или иной степенью уверенности считать символы на сравниваемых изображениях или совпадающими, или различными.

При чём предложенные меры отличия практически не зависят от контурных шумов сравниваемых символов. Рассмотрены семейства мер отличия, используемые в разных алгоритмах и алгоритмы разбиения на классы [3].

Для известных на сегодняшний день алгоритмов классификации, включая хорошо известный алгоритм JВ2, 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.

Приведен сравнительный анализ методов JPEG 2000, JBIG2, JВ2, ИЛЛ, сделан вывод о несомненном преимуществе алгоритма ИЛЛ: за счёт того, что словарь ИЛЛ получается почти в три раза короче, чем словарь JВ2, вытекает и превосходство в степени сжатия.

Сделаны выводы о том, что один из важнейших аспектов при обсуждении любого метода сжатия – качество восстановленного изображения по сравнению со 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. При классификации выделенных символов, если она обладает высоким качеством, получаются классы, состоящие из большого числа изображений одного и того же символа. Подходящая статистическая обработка этого класса позволяет избавиться от искажений, привнесенных при печати и сканировании и имеющих случайный характер [1].

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.

### Список литературы

1. Иванов В.Г., Ломоносов Ю.В., Любарский М.Г. *Сжатие изображения текста на основе выделения символов и их классификации*. Киев: *Международный научно-технический журнал «Проблемы управления и информатики»*, 2010, №6. С. 111 – 122.
2. *Прикладная статистика: Классификация и снижение размерности: [Справочник]* / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков и др.; Под ред. С.А. Айвазяна. – М.: *Финансы и статистика*, 1989. – 607с.
3. *Межиров И. Курсовая работа на тему «Алгоритмы сжатия данных»*. – Москва, МГУ им. Ломоносова, *механико-математический ф-т, научный руководитель А. Шень*, 2004.

**ТЕЗИ ДОПОВІДЕЙ**  
**VII міжнародної науково-практичної конференції**  
**“Проблеми та перспективи розвитку ІТ-індустрії”**  
**17 – 18 квітня 2015 р.**

Відповідальний за випуск: *С.П. Євсєєв*

Комп`ютерна верстка: *О.Г. Король*

---

Підписано до друку **29.11.2013**. Формат 60×84/8. Папір офсетний.  
Гарнітура «TimesNewRoman». Друк ризографічний. Ум.-друк. арк. – **9**. Ціна договірна.  
Наклад **250** прим.Зам. **1029-12**

---

Видавництво «Цифрова друкарня №1»  
Свідоцтво суб`єкта видавничої справи: серія ДК № 4354 від 06.07.2012 р.  
61001, м. Харків, пл. Повстання, 7/8  
e-mail: [zebra-zakaz@mail.ru](mailto:zebra-zakaz@mail.ru)

---

Віддруковано з готових оригінал-макетів у друкарні ФОП Петров В.В.  
Єдиний державний реєстр юридичних осіб та фізичних осіб-підприємців.  
Запис № 2480000000106167 від 08.01.2009.

61144, м. Харків, вул. Гв. Широнінців, 79в, к. 137, тел. **(057)778-60-34**  
e-mail: [bookfabric@rambler.ru](mailto:bookfabric@rambler.ru)