

## **FWD-формат электронных документов**

### **FDW-format of Electronic Documents**

## **FWD-формат електронних документів**

*Лигун А. А., Шумейко А. А., Савчук А. С.  
ООО «Альфа-Лоджик», Днепрпетровск, Украина*

*Павлюченков О. И.  
ООО «Техноинжиниринг», Днепрпетровск, Украина*

*A. A. Ligon, A. A. Shumeiko, and A. S. Savchuk  
Alfa-Logic Joint-stock Company, Dnepropetrovsk, Ukraine*

*O. I. Pavlyuchenkov  
Technoengineering Joint-stock Company, Dnepropetrovsk, Ukraine*

*Лігун А. А., Шумейко А. А., Савчук О. С.  
ТОВ «Альфа-Лоджик», Дніпрпетровськ, Україна*

*Павлюченков О. І.  
ТОВ «Техноінжиніринг», Дніпрпетровськ, Україна*

Описаны проблемы, возникающие при попытке реализации полнотекстового хранения и предоставления читателям библиотеки документов, и намечены пути решения этих проблем, реализованные в ALL-Library. Представлен разработанный ООО Альфа-Лоджик собственный формат данных АБИС (\*. fwd), полученный на основе современных технологий с использованием оригинальных методов сжатия документов различной структуры.

The problems of full-text document storage and lending are described and their solutions realized in the ALL-Library are outlined. The data format for ALIS designed by Alfa-Logic Company using the original methods of compression of the documents of various structures is presented.

У доповіді описані проблеми, що виникають при спробі реалізації повнотекстового зберігання і надання читачам бібліотеки документів, і намічені шляхи рішення цих проблем, реалізовані в ALL-Library. Представлений розроблений ТОВ Альфа-Лоджик власний формат даних АБИС (\*. fwd), одержаний на основі сучасних технологій з використанням оригінальних методів стиснення документів різної структури.

В последнее время Интернет становится наиболее эффективным каналом доставки разнообразной информации. Электронные массмедиа, газеты и журналы стали столь же привычными, как и традиционные печатные издания. Многие из них выходят в электронном виде даже раньше чем в печатном.

Этому способствуют технологии компьютерной верстки и печати. В частности с использованием стандарта PDF фирмы Adobe.

Но, тем не менее, достаточно много печатной продукции выпущенной в докомпьютерную эру остаются недоступными в электронном виде. Кроме того, существуют ограничения, обусловленные защитой авторских прав и политикой некоторых издателей, не позволяющие предоставлять документы в виде текста доступного для копирования и редактирования. Тоже относится и к рукописным источникам и историческим книгам, с пометками и оригинальными иллюстрациями.

Компанией «Техноинжиниринг» разработан графический формат FWD, предназначенный для хранения и размещения в сети Интернет электронных копий книг документов, рукописей, схем, технологических проспектов.

Особенностью этого формата является высокая степень сжатия (до 1000 раз) иллюстрированных текстовых документов.

В его основе формата лежат собственные адаптивные методы выделения и сжатия текстовых объектов, графических изображений и прочих элементов (логотипов, таблиц и т. п.)

Такой способ хранения и публикации библиотекой отсканированных документов учитывает и законодательные аспекты. Большинство издательств запрещает предоставления преобразованных в текстовый формат книг, но разрешает их ксерокопирование и перефотографирование по запросам читателей.

Подобные форматы созданы компаниями LizardTech<sup>1</sup>, США (DjVU) и Luratech<sup>2</sup>, Германия (LuraDocument, стандарт JPEG2000/Part6) и успешно внедряются в западных системах электронного документа оборота.

Основной метод DJVU, JPEG2000/Part 6 и LuraDocument является идея разделения документа на цветные слои (слой фона, слой иллюстраций, слой текстовых элементов)

В основе метода FWD лежит выделение символов и локализация элементов, не являющихся символами, то есть таблиц, рисунков и пр. Все, что не вошло в этот перечень есть фон документа. Предложенный подход вместе с разработанным нами методом сжатия иллюстраций FWT (Fast Wavelet Transformation) дает преимущество над описанными выше форматами от 30 до 50%

Изюминкой нашего формата является скорость работы и метод сжатия графических иллюстраций FWT. При равных условиях разработанный нами метод позволяет сохранить детали и качество, лучше чем существующие аналоги, прежде всего, DjVU и LuraDocument.

Обработка документов осуществляется с помощью программного обеспечения ALL-Documents (опция электронной библиотеки («ALL-Library»)).

После обработки и сжатия образов оригинального документа в полученном документе возможны организация навигации (на основе гиперссылок), а также указание его библиографических описаний облегчающие каталогизацию и поиск.

В настоящий момент нами продолжается работа над усовершенствованием формата FWD и ПО ALL-Documents для улучшения качества сжатия, и удобства пользователей.

На сегодняшний день доступны версии, как для традиционных персональных компьютеров, так и для карманных компьютеров на платформе Windows CE и PocketPC, а также plug-in для просмотра документов в стандартном окне интернет-браузера.

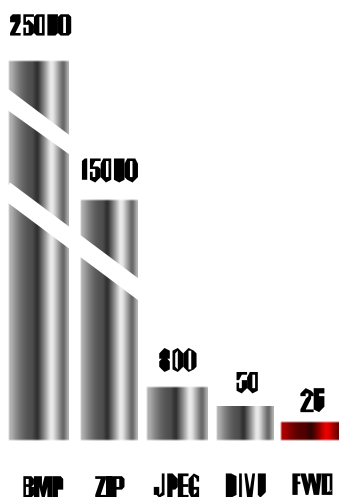


Рис. 1. Сравнение методов хранения графических изображений

<sup>1</sup> Веб-страница компании LizardTech <http://www.lizardtech.com>

<sup>2</sup> Веб-страница компании LuraTech <http://www.luratech.com>