**dedupass** - Многофункциональная высокопроизводительная утилита для комплексной

обработки, анализа и оптимизации словарей паролей. Обеспечивает продвинутую

дедупликацию с использованием хеш-алгоритмов, интеллектуальное слияние

множественных источников данных, пакетную обработку файловых масок и списков,

автоматическое определение кодировок и форматов. Поддерживает распределенную

обработку сверхбольших datasets (100+ ГБ), инкрементальное обновление словарей, сохранение метаданных и статистический анализ эффективности очистки. Включает

режимы пакетной командной строки, интерактивный wizard-интерфейс и API для

интеграции в пайплайны обработки данных безопасности.

*# Удалить дубликаты*

python passmerger.py --remove-duplicates rockyou.txt

*# Объединить файлы по маске*

python passmerger.py --output combined.txt \*.txt

*# Использовать файл-список*

python passmerger.py --list filelist.txt --output result.txt

*# Комбинированный подход*

python passmerger.py --list main.list --output mega.txt additional/\*.txt

**📋 Файлы-списки:**

# filelist.txt

\*.txt

@more\_files.list

d:\passwords\rockyou.txt

c:\data\\*.txt

**🖥️ Интерфейс:**

* Удаление дубликатов
* Объединение файлов через маски
* Работа с файлами-списками
* Поддержка рекурсивных масок

# **DEDUPASS** - Справка по использованию утилиты

## 📖 ОПИСАНИЕ

DEDUPASS - мощная утилита для работы со словарями паролей. Поддерживает удаление дубликатов, объединение файлов, работу с масками и файлами-списками. Оптимизирована для работы с файлами любого размера (до сотен GB).

## 🚀 КОМАНДЫ

### 1. Удаление дубликатов в файле

*# Удалить дубликаты в файле (изменяет исходный файл)*

dedupass.exe --remove-duplicates файл.txt

dedupass.exe -rd файл.txt

### 2. Создание очищенной копии

*# Создать очищенную копию файла (исходный файл не изменяется)*

dedupass.exe --output результат.txt исходник.txt

dedupass.exe -o чистый.txt грязный.txt

### 3. Объединение нескольких файлов

*# Объединить несколько файлов с удалением дубликатов*

dedupass.exe --output результат.txt файл1.txt файл2.txt файл3.txt

dedupass.exe -o общий.txt папка/\*.txt

### 4. Использование файла-списка

bash

*# Использовать файл со списком файлов для обработки*

dedupass.exe --list список.txt --output результат.txt

dedupass.exe -l файлы.list -o итог.txt

### 5. Комбинирование источников

*# Комбинировать файл-список и маски*

dedupass.exe --list основной.list --output mega.txt доп/\*.txt

### 6. Показать справку

*# Показать подробную справку*

dedupass.exe --help

dedupass.exe -h

## 🎯 ПРАКТИЧЕСКИЕ ПРИМЕРЫ

### Пример 1: Очистка скачанного словаря

*# Очистить файл от дубликатов*

dedupass.exe --remove-duplicates rockyou.txt

*# Создать очищенную копию*

dedupass.exe --output clean\_rockyou.txt rockyou.txt

### Пример 2: Объединение нескольких словарей

*# Объединить все txt файлы в папке*

dedupass.exe --output combined.txt passwords/\*.txt

*# Объединить конкретные файлы*

dedupass.exe --output mega\_passwords.txt rockyou.txt common.txt weak.txt

### Пример 3: Работа с большими наборами данных

*# Использовать файл-список для сложных сценариев*

dedupass.exe --list all\_files.list --output complete\_list.txt

*# Рекурсивная обработка*

dedupass.exe --output all.txt \*\*/\*.txt

## 📁 ФАЙЛЫ-СПИСКИ

### Формат файла-списка:

# Файл со списком файлов для обработки

# Комментарии начинаются с решетки

# Отдельные файлы

C:\passwords\rockyou.txt

/home/user/wordlists/common.txt

# Маски файлов

passwords/\*.txt

data/2024\_\*.txt

# Рекурсивные маски

wordlists/\*\*/\*.txt

# Вложенные списки

@more\_files.list

# Дополнительные файлы

extra\_passwords.txt

### Пример файла **my\_files.list**:

text

# Основные словари

rockyou.txt

common\_passwords.txt

# Файлы из папки

wordlists/\*.txt

# Дополнительные данные

@additional\_files.list

## 🎪 МАСКИ ФАЙЛОВ

### Базовые маски:

* \*.txt - все txt файлы в текущей папке
* passwords/\*.txt - все txt файлы в папке passwords
* data/2024\_\*.txt - файлы начинающиеся с "2024\_" в папке data

### Рекурсивные маски:

* \*\*/\*.txt - все txt файлы в текущей папке и подпапках
* wordlists/\*\*/\*.txt - все txt файлы в папке wordlists и ее подпапках

### Специальные символы:

* \* - любая последовательность символов
* ? - любой один символ
* [abc] - любой из символов a, b, c

## 📊 ОСОБЕННОСТИ РАБОТЫ

### Автоматическая оптимизация:

* Автоматически выбирает метод обработки в зависимости от размера файла
* Для маленьких файлов (<100MB) - быстрый метод в памяти
* Для больших файлов (>100MB) - потоковая обработка с чанками

### Визуальная обратная связь:

* Показывает размер файлов до и после обработки
* Отображает количество удаленных дубликатов
* Приводит примеры удаленных строк для проверки

### Безопасность:

* Исходные файлы не изменяются при использовании --output
* Сохраняется порядок строк при обработке
* Поддержка разных кодировок и форматов переводов строк

## 🔧 ТЕХНИЧЕСКИЕ ДЕТАЛИ

### Поддерживаемые размеры файлов:

* ✅ Малые файлы: < 100 MB (быстрая обработка в RAM)
* ✅ Средние файлы: 100 MB - 500 MB (оптимизированная обработка)
* ✅ Большие файлы: 500 MB - 100+ GB (потоковая обработка)

### Форматы обработки:

* Удаление пустых строк
* Обрезка пробелов по краям
* Унификация переводов строк (всегда \n)
* Удаление BOM маркеров кодировки

### Алгоритмы:

* MD5 хеширование для экономии памяти
* Внешняя сортировка для больших файлов
* Многоуровневое удаление дубликатов

## 💡 СОВЕТЫ ПО ИСПОЛЬЗОВАНИЮ

### Для максимальной производительности:

bash

*# Сначала очистить отдельные файлы*

dedupass.exe -rd file1.txt

dedupass.exe -rd file2.txt

*# Потом объединить*

dedupass.exe -o result.txt file1.txt file2.txt

### Для работы с огромными файлами:

bash

*# Использовать файлы-списки для организации*

dedupass.exe -l huge\_files.list -o result.txt

*# Рекурсивный сбор файлов*

dedupass.exe -o all.txt \*\*/\*.txt

### Для отладки и проверки:

bash

*# Посмотреть примеры удаленных дубликатов*

dedupass.exe -rd file.txt

*# Проверить работу на маленьком файле*

dedupass.exe -o test\_result.txt small\_file.txt

## 🆘 ПОЛУЧЕНИЕ ПОМОЩИ

### Интерактивный режим:

bash

*# Запустить интерактивное меню*

dedupass.exe

### Командная справка:

bash

*# Краткая справка*

dedupass.exe --help

*# Подробная справка*

dedupass.exe -h

## 🐛 РЕШЕНИЕ ПРОБЛЕМ

### Файл не найден:

* Проверьте правильность пути
* Используйте абсолютные пути для файлов в других папках
* Убедитесь, что файл существует

### Большой размер итогового файла:

* Программа предлагает оптимизацию при значительном увеличении размера
* Можно подтвердить оптимизацию

### Проблемы с кодировкой:

* Программа автоматически обрабатывает разные кодировки
* Используется обработка ошибок 'ignore' для проблемных символов

**Связь**:

О выявленных ошибках и предложения пишите:

* Telegram: <https://t.me/avhelpnew>
* [artemsoft@yahoo.com](mailto:artemsoft@yahoo.com)