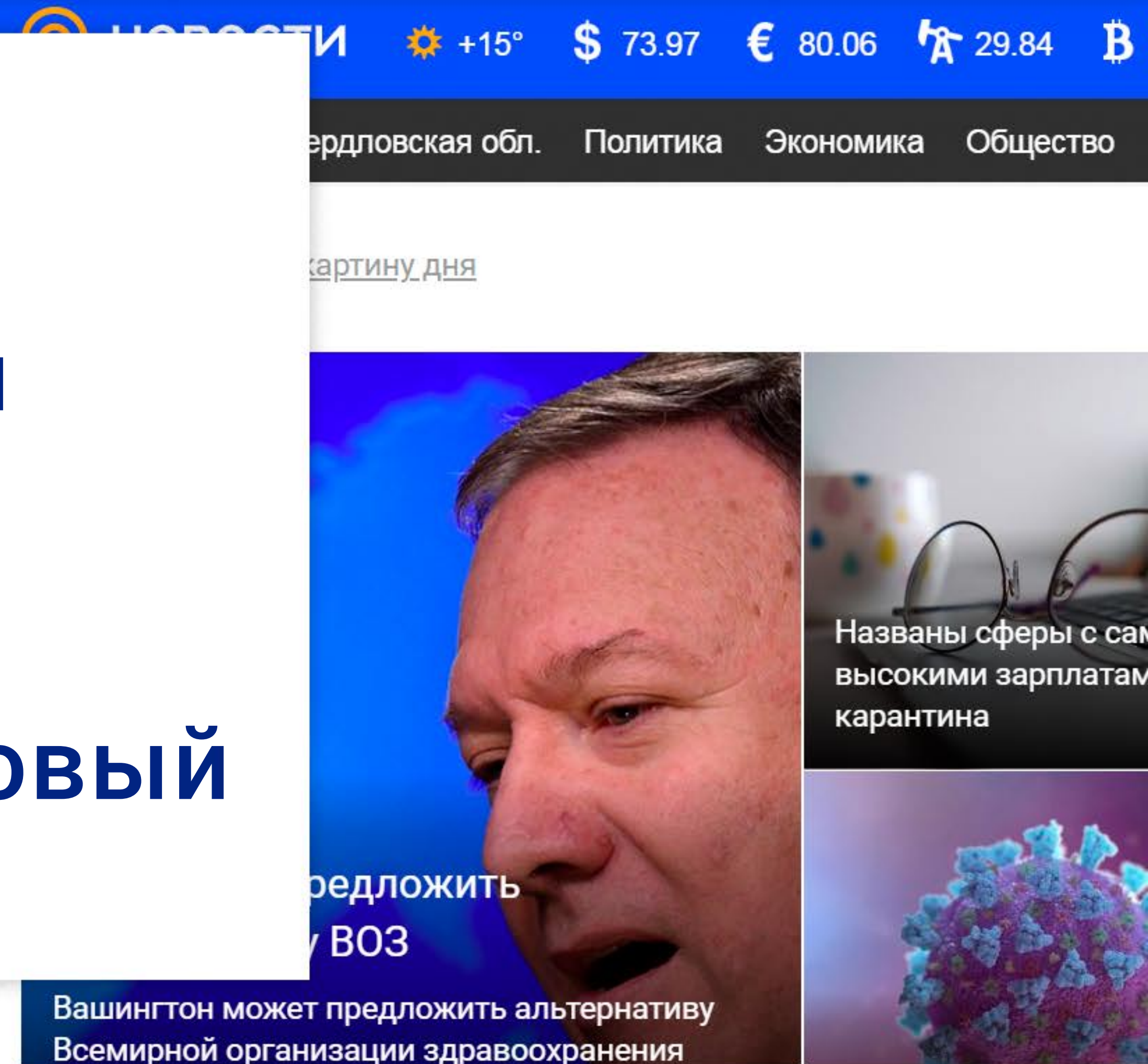


# Парсер контента с раздела новости mail.ru

## Алгоритм работы:

- ищет страницу по ссылке;
- забирает заголовок, дату и текст новости;
- чистит разметку;
- сохраняет контент в текстовый файл.





# Получение данных

Для взаимодействия с сервером программа использует ресурсы библиотеки libcurl.

Это позволяет:

- точно настраивать параметры запроса;
- модернизировать функционал.



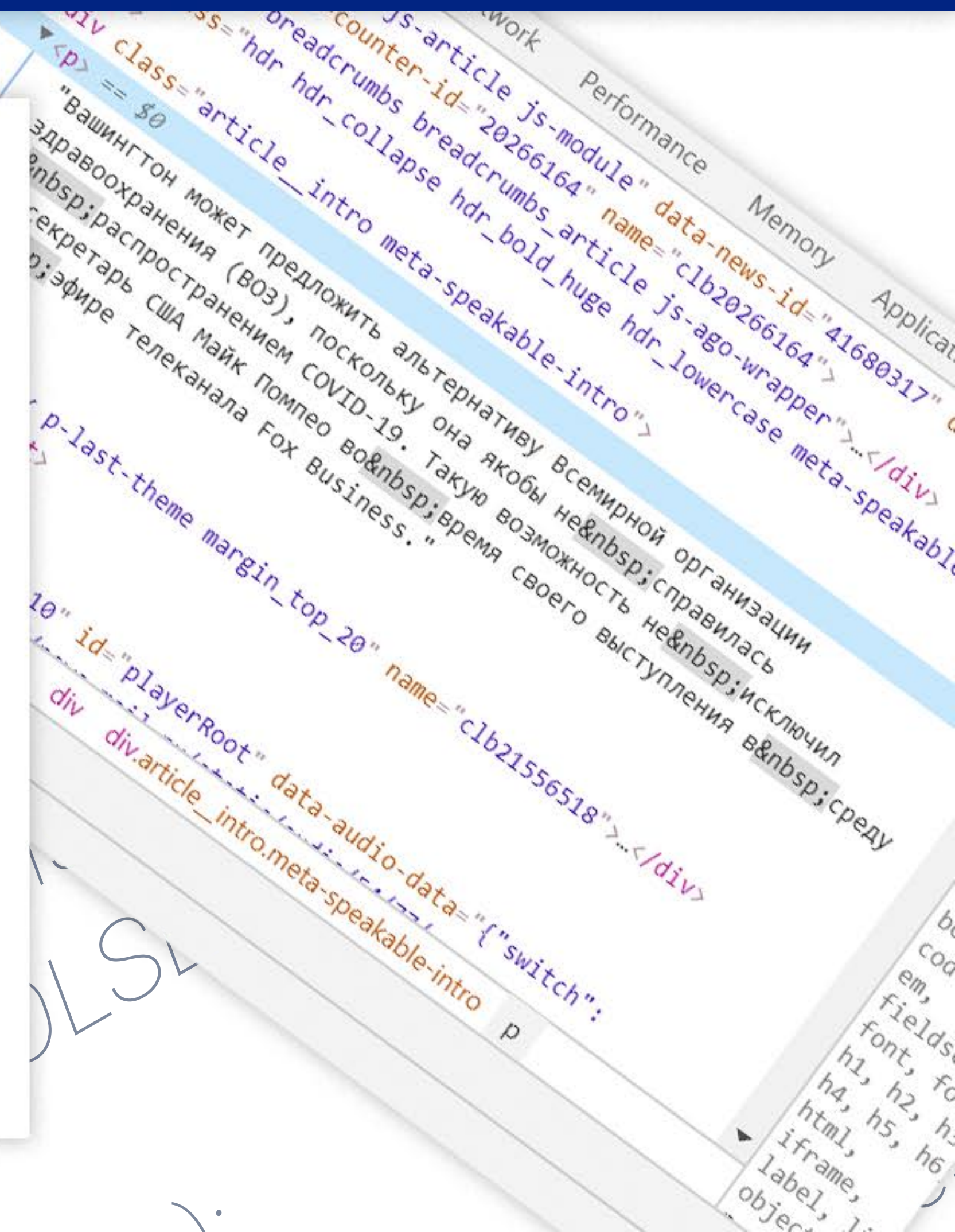


# Поиск и фильтрация контента

**Поиск контента в html-разметке осуществляется:**

- по тегу h1 для заголовка;**
- по классу datetime для даты;**
- по классу article + p для текста.**

**Дополнительно программа чистит ссылки и запреты на перенос.**





# Удобство использования

**Исходный код программы с инструкцией размещен на сервисе GitHub по адресу:**

[https://github.com/Alexcppio/MailRuNewsParser.](https://github.com/Alexcppio/MailRuNewsParser)

**Архив с инсталлятором дает возможность установить парсер на любой компьютер и сразу приступить к использованию.**

Программа записывает контент в текстовый файл с расширением txt, что позволяет сразу получать данные, без какого-либо форматирования.





# Варианты модернизации ПО

**Расширение функционала приложения в перспективе может включать:**

- автоматизация получения новостей, с определением категории и свежести;**
- реализация доступа по логину-паролю;**
- подключение модуля поставки свежих прокси и доступ через прокси;**
- вынесение в отдельный модуль словаря лексем для выгрузки контента со всех популярных новостных порталов.**

