

## **Abstrak**

**Informasi merupakan salah satu hal yang penting pada era sekarang, salah satu informasi yang selalu update setiap harinya adalah berita. Banyaknya berita-berita yang muncul setiap harinya menjadi sebuah masalah baru ketika website berita tidak menyediakan layanan API (Application Programming Interface) untuk mendapatkan berita-berita tersebut. Hal tersebut menjadi kendala bagi peneliti yang akan melakukan analisis topik berita. Cara salin dan tempel kurang efektif digunakan untuk mendapatkan berita setiap harinya pada website berita dikarenakan membutuhkan waktu yang cukup lama. Pada tugas akhir ini dilakukan web scraping dengan metode HTML (HyperText Markup Language) DOM (Document Object Model) untuk ekstraksi data dari situs berita, hasil web scraping berupa dataset yang kemudian dijadikan sebuah API. API yang telah dibuat akan diuji kesesuaian datanya, antara data yang diperoleh pada saat melakukan scraping dan data yang ada pada website berita pada saat dilakukan pengujian. Hasil pengujian menunjukkan bahwa website berita Tribunnews.com memiliki tingkat kesesuaian sebesar 99,2%, Detik.com sebesar 97,9% dan Liputan6.com sebesar 98,6%.**

**Kata kunci : Web Scraping, HTML DOM, Situs Web, Berita, API.**