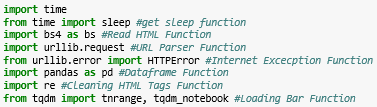
Dokumentasi Tribunnews Crawler

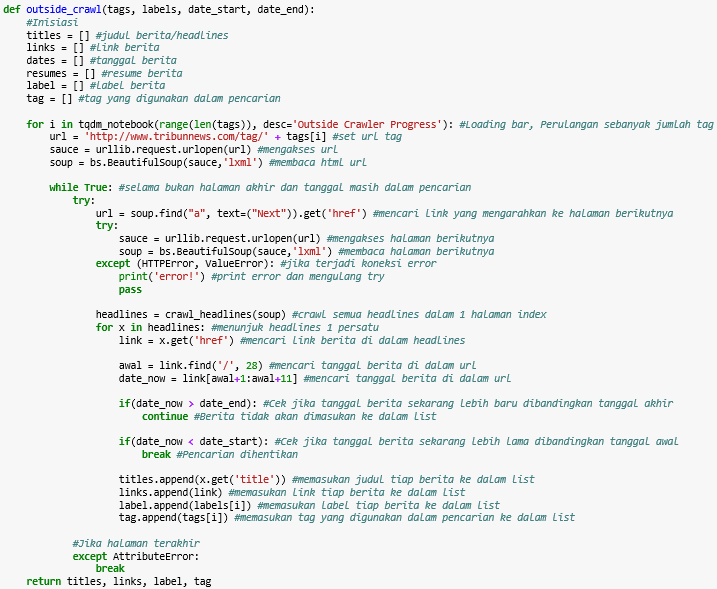
Source Code



Library yang digunakan



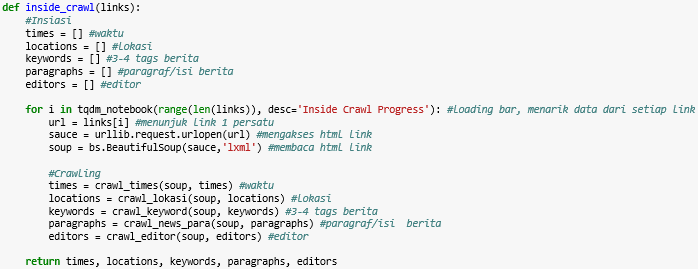
Untuk Menghapus Tags HTML seperti <p>



Crawl Halaman Index Berita



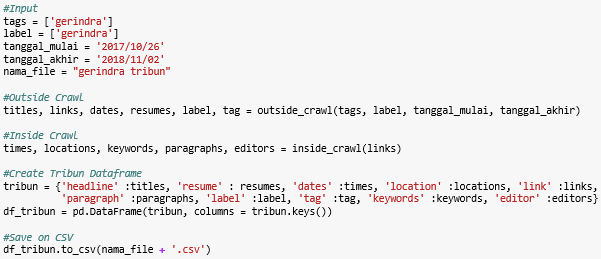
Tampilan Halaman Index Berita



Crawl Bagian dalam Berita



Tampilan Bagian Dalam Berita



Fungsi Utama Crawler

CSV

Headline: Headline dari berita

Dates: Tanggal dan Waktu dari berita

Location: Lokasi dari berita

Link: URL berita

Paragraph: Paragraf dari berita

Label: Laber dari input

Tag: Tags dari input

Keywords: 3 Tag dari berita

Editor: Editor dari berita



Headline

Tags



Crawl Headlines

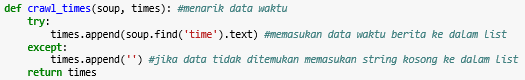


Link

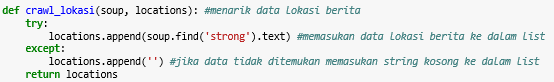
Date

Paragraph

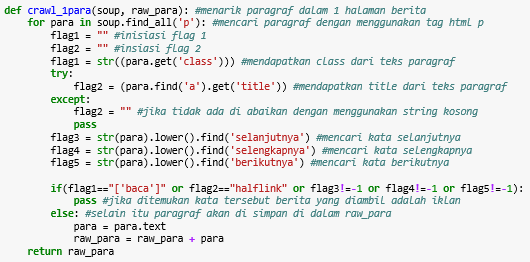
Location



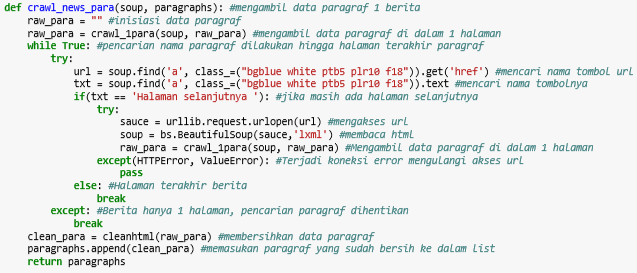
Crawl Date



Crawl Location



Crawl Paragraf dalam 1 halaman



Crawl Paragraf seisi berita

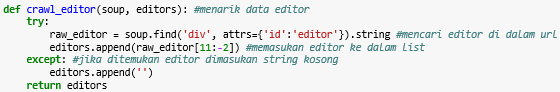


Editor

Keywords



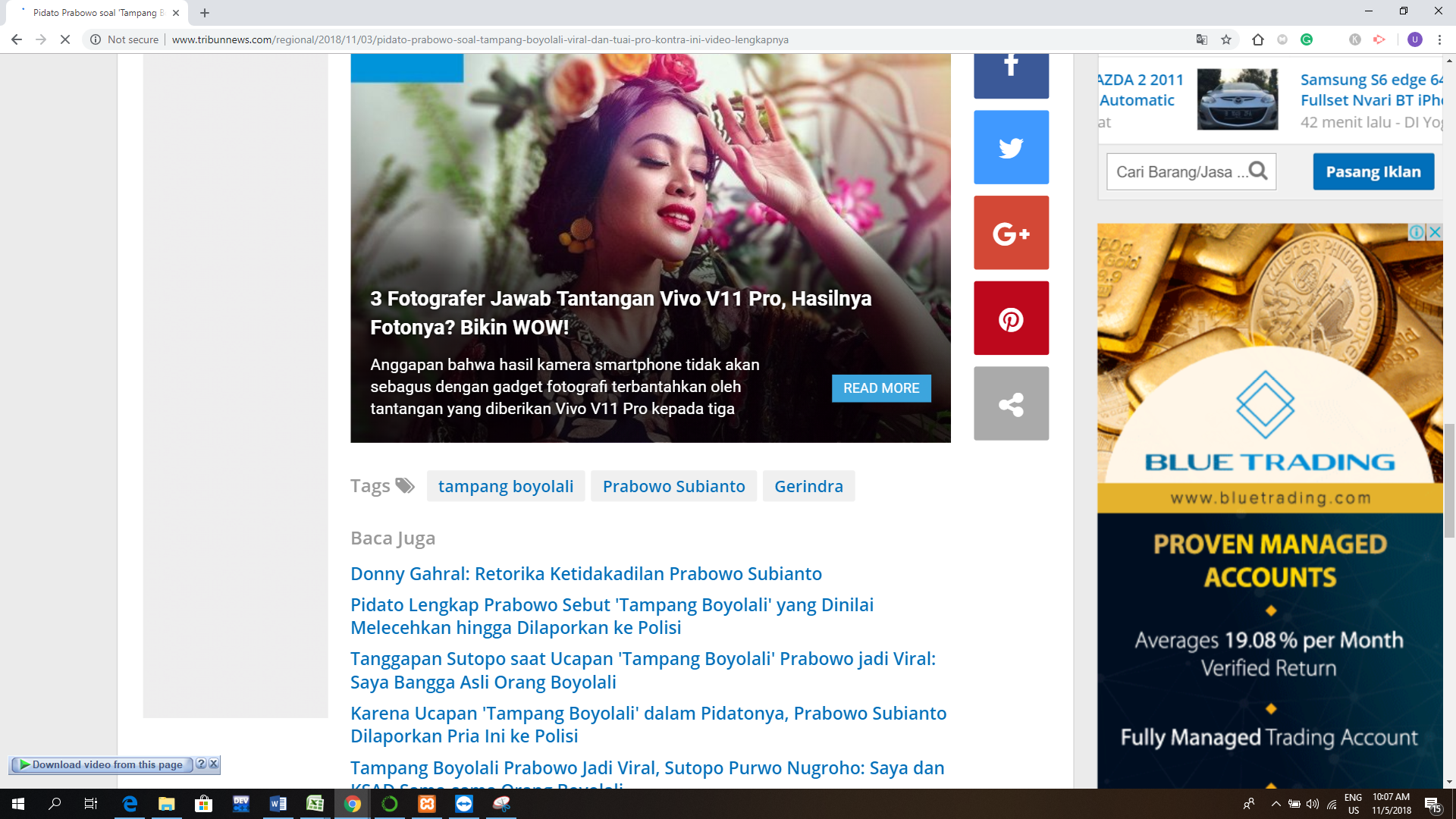
Crawl Keywords



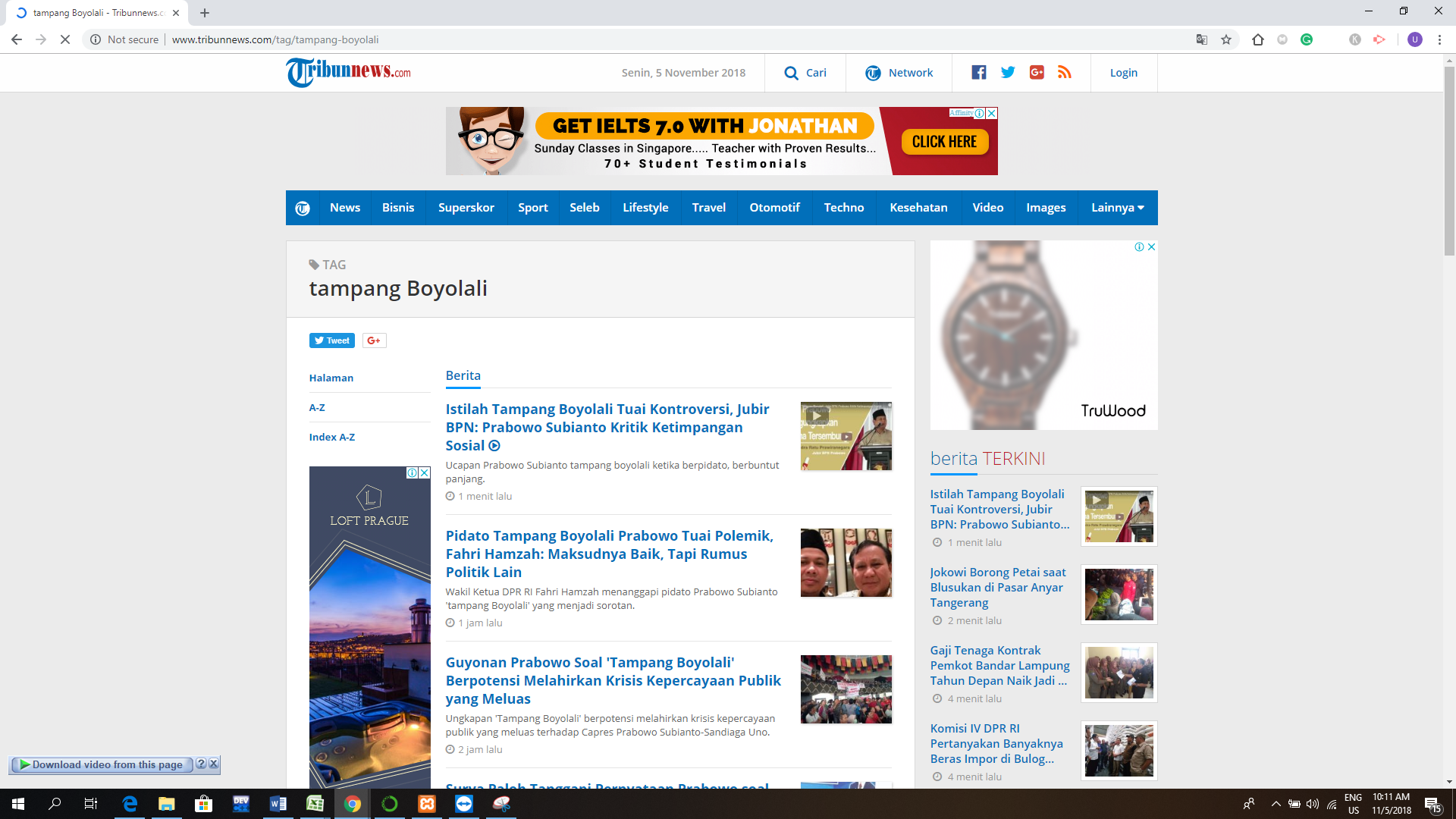
Crawl Editor

Notes

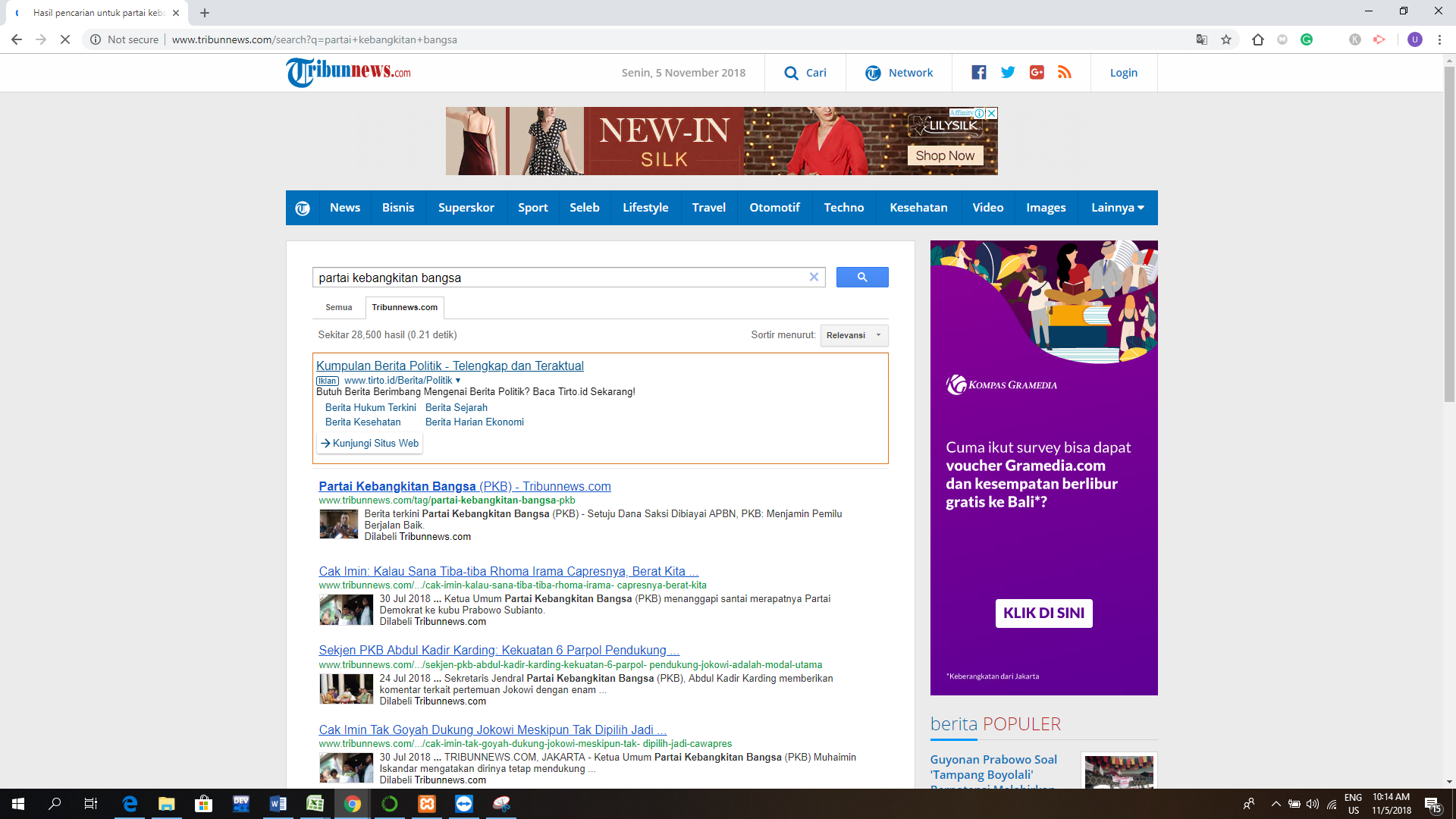
Cara memastikan tags



Jika Tags lebih dari satu kata disambungkan “-“. Cth: tampang-boyolali, Prabowo-Subianto.



Untuk Cek Tags dapat dengan mengakses URL: [http://www.tribunnews.com/tag/<tags](http://www.tribunnews.com/tag/%3ctags)>. Cth: <http://www.tribunnews.com/tag/tampang-boyolali>. Jika Merupakan Tags akan menampilkan seperti gambar di atas.

Jika bukan merupakan tags tampilan akan berubah menjadi seperti gambar di atas. Selain itu, URL akan berubah menjadi <http://www.tribunnews.com/search>?.