

## PERBANDINGAN KINERJA ALGORITMA STRING MATCHING BOYER-MOORE & KNUTH-MORRIS-PRATT PADA SEO WEB SERVER

Sena Ramadona Cakrawijaya<sup>1,\*</sup>, Bambang Kriswantara<sup>2)</sup>

<sup>1, 2)</sup>Magister of Computer Science, Faculty of Engineering, STMIK Nusa Mandiri  
Jalan Kramat Raya No. 18, Senen, Jakarta Pusat, Indonesia

\*Corresponding Author: 14002412@nusamandiri.ac.id

### Abstrak

Algoritma pencocokan string mencoba menemukan posisi di mana satu atau beberapa pola (juga disebut string) dapat terjadi dalam teks. Penelitian ini membandingkan algoritma Knuth-Morris-Pratt dan Boyer-Moore untuk dapat mencocokkan algoritma di internet atau dokumen web. Pada internet atau dokumen web, pencarian adalah proses penting untuk melakukan ekstraksi konten. Oleh karena itu, panjang tag HTML diperiksa agar dapat menentukan algoritma mana yang lebih cocok untuk proses pencocokan. Eksperimen menunjukkan bahwa algoritma yang dipilih adalah algoritma pencocokan "pola terbaik" dengan kinerja 12 kali lipat lebih cepat dibandingkan yang algoritma lain dalam worst-case untuk SEO HTML tags dokumen server web. Hasil tersebut berbeda dengan literatur sebelumnya yang menunjukkan kinerja sebaliknya.

**Kata kunci:** algoritma string matching; knuth-morris-pratt; boyer-moore; web server; seo

### Abstract

String matching algorithms try to find positions where one or more patterns (also called strings) occur in the text. This study compares Knuth Morris Pratt and Boyer Moore algorithm for matching algorithms on the internet or web documents. On the Internet or web documents, searching is a crucial process for the content extraction process. Therefore, the lengths of HTML tags are examined for determining which algorithm is more suitable for the matching process. This experiments show that the chosen algorithm is the best pattern matching algorithm with more than 12 times faster than the other in the worst-case scenario for SEO HTML tags web server documents. These results are different from previous literature that showed the opposite performance.

**Keywords:** string matching algorithm; knuth-morris-pratt; boyer-moore; web server; seo

### 1. Pendahuluan

Teknik pencarian *string matching* yang sudah populer menggunakan Knuth Morris Pratt (KMP) dan Boyer Moore (BM) ditemukan pada tahun yang sama, yaitu pada tahun 1977 [5][6]. Algoritma KMP dan BM dibuat untuk pencocokan *string* yang efektif dalam tujuan pencarian. Meskipun lebih dari 30 tahun algoritma ini sudah ada, masih banyak aplikasi mengadopsinya untuk teknik pencarian. Saat ini fitur pencarian informasi berbasis web adalah suatu keharusan, untuk alasan itu algoritma pencarian yang lebih baik antara KMP dan BM perlu dipilih agar aplikasi pada server web memiliki kinerja pencarian yang optimal.

Algoritma Boyer Moore diperkenalkan oleh Robert S. Boyer dan J. Strother Moore pada tahun 1977. Pencocokan karakter yang dibuat menggunakan algoritma Boyer-Moore dimulai dari string dengan "pola kanan". Ide di balik algoritma ini adalah untuk memulai mencocokkan karakter dari kanan dan kemudian informasi lebih lanjut dapat diperoleh [2].

Algoritma Knuth Morris Pratt (KMP) adalah algoritma yang digunakan untuk melakukan proses pencocokan string. Algoritma ini adalah jenis algoritma pencocokan string sama persis yang merupakan string yang tepat dan cocok dengan susunan karakter dalam string yang sama yang memiliki jumlah serta urutan karakter dalam string yang sama pula [3]. Contoh kata algoritmik hanya menampilkan kecocokan dengan kata algoritmik. Dalam algoritma KMP,

informasi yang digunakan untuk melakukan pergeseran lebih lanjut, bukan hanya satu karakter seperti algoritma Brute Force. Algoritma ini melakukan pencocokan dari kiri ke kanan.

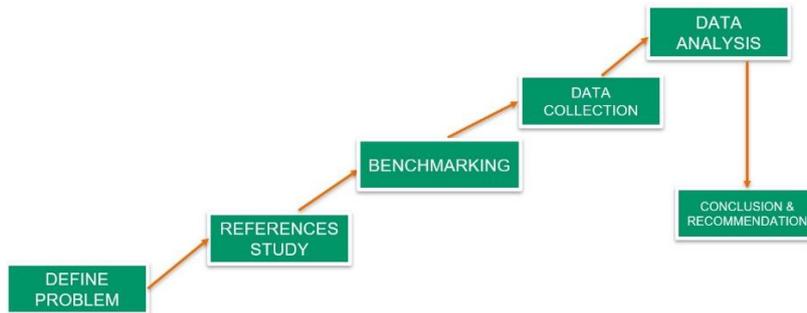
Kedua algoritma dibandingkan karena teknik kesamaan *string* yang cocok untuk pencarian, terutama untuk pencarian string pada dokumen file teks. Dalam aplikasi berbasis web, informasi utama ditulis dalam *format hypertext markup language* (HTML), karena fitur pencarian harus secara efisien menemukan teks *string* di web yang memenuhi kondisi itu.

*Search engine optimization* (SEO) sering digunakan untuk meningkatkan peringkat situs web agar lebih banyak terekspos ke dunia online (internet). *Tag* HTML dapat digunakan untuk membuat teknik SEO yang lebih baik. Karena itu, pencarian pola dalam *tag* HTML menjadi penting untuk dilakukan. Ada beberapa poin SEO dalam *tag* HTML [8] seperti judul, deskripsi *meta*, dan judul. Itu hanya 3 teratas SEO yang dapat diuji dalam penelitian ini. Semua penelitian diuji dalam skenario nyata seperti pencarian web server secara langsung.

Penelitian awal tentang kinerja KMP dan BM telah diterbitkan dan berhasil menunjukkan perbandingan relatif antara perbandingan kinerjanya [1]. Dalam penelitian ini dilakukan uji perbandingan dalam perbandingan kinerja mendalam dengan berbagai tolak ukur dan pengujian file yang lebih besar sehingga hasilnya bisa jauh lebih baik untuk menunjukkan apa yang lebih unggul untuk studi kasus dari analisis komparatif. Pentingnya penelitian ini adalah untuk menentukan algoritma mana yang lebih cocok untuk digunakan di Web Server sebagai fitur pencariannya.

## 2. Metode Penelitian

Proses metode penelitian diberikan pada Gambar 1. Dimulai dari mendefinisikan masalah termasuk menentukan ruang lingkungnya.



Gambar 1. Metodologi Penelitian

Masalah yang ada didefinisikan untuk mengumpulkan referensi yang memadai, setelah semua referensi dikumpulkan, implementasi tolak ukur dimulai. Status pengujian selesai setelah setiap variabel pengujian telah lulus. Pengumpulan data dan analisis data dilakukan setelah itu. Langkah terakhir adalah membuat kesimpulan dan rekomendasi algoritma mana yang lebih cocok untuk kasus ini.

### 2.1 Define Problem

Langkah pertama untuk memecahkan masalah dengan model analitik adalah mendefinisikan masalah dengan tepat, mengidentifikasi masalah yang ada sehingga dapat ditentukan ruang lingkup penelitian yang harus dilakukan dan arah solusi untuk diperoleh hasil yang lebih baik.

### 2.2 References Study

Studi literatur yang ada yang ditemukan mengenai judul dan studi kasus yang harus dilakukan dalam penelitian ini. Jurnal referensi sangat penting untuk menentukan titik awal. Ada penelitian di masa depan yang harus dilakukan dari referensi dan mencoba untuk memecahkan dalam penelitian ini.

### 2.3 Benchmarking

*Benchmarking* adalah proses pengukuran kinerja algoritma. Inti dari benchmarking adalah untuk mengidentifikasi peluang atau potensi untuk perbaikan [9]. Mempelajari algoritma dengan kinerja yang unggul, memecahkan apa yang membuat kinerja unggul seperti itu memungkinkan dan kemudian membandingkan proses tersebut dengan bagaimana algoritma beroperasi, dapat mengimplementasikan perubahan yang menghasilkan peningkatan yang signifikan. Pada tahap ini menggunakan Bahasa Indonesia sebagai alat benchmarking dan menyiapkan hardware (hardware) dan perangkat lunak pendukung sehingga proses pengujian dapat terlaksana dengan baik.

### 2.4 Data Collection

Proses pengumpulan dan pengukuran informasi tentang variabel yang diminati, dengan cara sistematis yang robust memungkinkan seseorang untuk menjawab pertanyaan penelitian yang dinyatakan, menguji hipotesis, dan mengevaluasi hasil. Komponen pengumpulan data penelitian adalah umum untuk semua bidang studi. Sementara metode bervariasi menurut disiplin ilmu, memastikan pengumpulan data yang akurat dan jujur. Mengumpulkan data output dan menghitung nilai rata-rata setiap algoritma sehingga dapat disajikan informatif dalam bentuk tabel dan gambar.

### 2.5 Data Analysis

Analisis data adalah proses memeriksa, membersihkan, mengubah, dan memodelkan data dengan tujuan menemukan informasi yang berguna, menginformasikan kesimpulan dan mendukung pengambilan keputusan [11]. Analisis data memiliki beberapa aspek dan pendekatan, mencakup berbagai teknik. Analisis data output algoritma mana yang terbaik untuk kasus ini termasuk juga menyajikan rekapitulasi data sehingga hasil tes dapat berdampak signifikan pada hasil keseluruhan penelitian.

### 2.6 Conclusion & Recommendation

Validitas adalah tingkat kesimpulan dimana adanya hubungan diantara variabel berdasarkan data yang benar atau "masuk akal" [12]. Pada tahap akhir setelah rangkaian langkah dan tes, tahap terakhir adalah membuat kesimpulan dari penelitian ini, algoritma terbaik untuk kasus ini harus sampai pada kesimpulan yang menunjukkan berapa banyak perbedaan dengan algoritma lainnya dan menyatakan hasil kajian serta penelitian lebih lanjut yang dapat dilanjutkan dengan memberikan beberapa rekomendasi terkait penelitian yang telah dilakukan.

## 3. Hasil dan Pembahasan

Pada Gambar 2 dan 3 ada file HTML yang dilakukan uji pencarian dengan "pencarian pola string" dengan mengingat dan membandingkan dua string dilakukan adanya ketidakcocokan string atau kata [13]. *Website 1* adalah file HTML dari situs berita internasional yang telah diindeks peringkat Google sebagai Peringkat teratas [14], sedangkan *Website 2* adalah file HTML situs berita lokal di Indonesia juga dengan peringkat halaman teratas Google [14].

Salah satu yang mencari tahap ini adalah judul tag HTML dari setiap file, ditandai dengan tag judul <title>. Kedua adalah tag <meta>meta yang biasanya berisi kata kunci konten halaman situs web, yang terakhir adalah tag <h1>judul yang berisi konten utama halaman situs web berbasis HTML [8].

Dilakukan 6 pencarian dengan kata kunci atau kalimat yang berbeda untuk algoritma KMP dan BM terkandung dalam Tabel 1 sehingga perbedaan kinerja dapat dilihat secara komprehensif karena menggunakan kata kunci pada perangkat komputer yang sama dalam perangkat lunak dan perangkat kerasnya.

Tabel 1. Kata kunci

	KMP	BM
Keyword-1-HTML-1	top secret	top secret
Keyword-2-HTML-1	ufo experts have long suspected the government	ufo experts have long suspected the government
Keyword-3-HTML-1	extraterrestrial phenomena	extraterrestrial phenomena
Keyword-1-HTML-2	video youtube paling trending	video youtube paling trending
Keyword-2-HTML-2	paling menarik perhatian masyarakat indonesia	paling menarik perhatian masyarakat indonesia
Keyword-3-HTML-2	rahasia tetangga	rahasia tetangga

Pada Tabel 1 juga dilakukan pengujian pencarian menggunakan Bahasa Indonesia dan Bahasa Inggris untuk mengetahui apakah ada perbedaan kinerja diantara keduanya. Jumlah kata yang dimasukan bervariasi antara 2 sampai dengan 7 kata dalam satu kalimat. Contoh pencarian mengacu pada topik *trending* yang paling banyak dicari pada saat ini [15].

Isi HTML suatu website yang dilakukan pencarian kata atau kalimat seperti yang ditunjukkan pada Gambar 2 dan 3. Pencarian dilakukan dengan urutan teratas sampai terbawah dari tiap barisnya.

```

1 <title>UFOs: Top Secret Evidence Revealed | HISTORY</title>
2 <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
3 <meta name="description" content="UFO experts have long suspected the government of
4 <meta name="viewport" content="width=device-width, initial-scale=1.0">
5 <meta name="format-detection" content="telephone=no">
6 <meta name="user-agent" content="IE 11.0.0 / Windows 7.0.0">
7 <meta name="device" content="desktop">
8 <meta name="robots" content="index, follow">
9 <meta name="a:cn:countryCode" content="US">
10 <meta name="a:cn:device" content="desktop">
11 <meta name="a:cn:platform" content="webcenter">
12 <meta name="a:cn:env" content="prod">
13 <meta name="a:cn:a:cn:prod" content="prod">
14 <meta name="a:cn:brandKey" content="history">
15 <meta name="a:cn:subbrand" content="HISTORY">
16 <meta name="a:cn:template" content="special-landing">
17 <meta name="a:cn:contentType" content="special">
18 <meta name="a:cn:specialId" content="215571">
19 <meta name="a:cn:specialTitle" content="UFOs: Top Secret Evidence Revealed">
20 <meta property="fb:app_id" content="125536057474924">
21 <meta property="fb:admins" content="100001148659782">
22 <meta property="og:site_name" content="HISTORY">
23 <meta property="og:type" content="article">
24 <meta property="og:title" content="UFOs: Top Secret Evidence Revealed | HISTORY">
25 <meta property="og:description" content="UFO experts have long suspected the govern
26 <meta property="og:url" content="https://www.history.com/specials/ufo-top-secret-e
27 <meta property="og:image" content="">
28 <meta name="twitter:card" content="summary_large_image">
29 <meta name="twitter:url" content="https://www.history.com/specials/ufo-top-secret-
30 <meta name="twitter:title" content="UFOs: Top Secret Evidence Revealed | HISTORY">
31 <meta name="twitter:description" content="UFO experts have long suspected the gover
32 <meta name="twitter:site" content="@history">
33 <meta name="twitter:image" content="">
    
```

Gambar 2. HTML Website-1

```

1 <html lang="id"><head> YouTube
2 <meta charset="utf-8">
3 <meta http-equiv="X-UA-Compatible" content="IE=edge">
4 <meta name="viewport" content="initial-scale=1.0, user-scalable=no, width=device-width, height
5 <meta name="apple-mobile-web-app-capable" content="yes">
6 <meta name="application-filecolor" content="#3C214D">
7 <link rel="shortcut icon" href="https://cdn.detik.net.id/inet2/images/favicon.ico?w=8a5d0e0a">
8 <link rel="apple-touch-icon" href="https://cdn.detik.net.id/inet2/images/favicon.ico?w=8a5d0e0a">
9 <meta name="theme-color" content="#20409A">
10 <meta name="google-site-verification" content="Iqj4hy17QfCai3UoWhumClTACR61J_WWfyYVFY83HA">
11 <meta content="id" name="geo.country">
12 <meta http-equiv="content-language" content="In-Id">
13
14 <title>Deretan Video YouTube Paling Trending di 2019</title>
15 <meta name="description" content="Setiap akhir tahun YouTube selalu merilis dereta
16 <meta charset="utf-8">
17 <meta property="og:type" content="article">
18 <meta property="og:site_name" content="detikinet">
19 <meta property="og:title" content="Deretan Video YouTube Paling Trending di 2019">
20 <meta property="og:image" content="https://akcdn.detik.net.id/api/wm/2019/10/17/0053280a-b
21 <meta property="og:description" content="Setiap akhir tahun YouTube selalu merilis deretan
22 <meta property="og:url" content="https://inet.detik.com/cyberlife/d-4812201/deretan-video-
23 <meta property="fb:app_id" content="187960271237149">
24 <meta property="fb:admins" content="10000607566694">
25 <meta property="og:image:type" content="image/jpeg">
26 <meta property="og:image:width" content="650">
27 <meta property="og:image:height" content="366">
28 <meta name="viewport" content="width=device-width,minimum-scale=1,initial-scale=1">
29 <link href="https://plus.google.com/detikcom" rel="publisher">
30 <meta name="copyright" content="detikinet" itemprop="datePublished">
31 <meta name="p:domain-verify" content="2057b86bf61e5a346e22a380cfecf89"><meta name="robots
32 <meta name="articleid" content="4812201">
33 <meta name="articletype" content="singlepage">
34 <meta name="articledewasa" content="dewasitidak">
35 <meta name="articlehow" content="default">
    
```

Gambar 3. HTML Website-2

Gambar 2 dan 3 menunjukkan kata atau kalimat yang dicari berada diantara tag HTML penanda kode tiap barisnya. Kata atau kalimat yang tampil dalam suatu website biasanya berada pada tag HTML "content" sehingga dapat mudah dilakukan pengamatan pada HTML seperti yang tampil pada halaman websitenya.

Setelah melakukan *benchmarking*, hasil dari seluruh output data dikumpulkan dan dilakukan rekapitulasi sehingga dapat ditampilkan data hasil sesuai Tabel 2. Hasil pencarian kata kunci dapat diselesaikan kecuali kasus ke-6 karena tidak ada kata kunci seperti itu dalam file HTML yang dicari.

Tabel 2. Hasil Algoritma

	KMP				BM	
	Kasus 1(s)	Kasus 2(s)	Kasus 3(s)	Kasus 4(s)	Kasus 5(s)	Kasus 6(s)
HTML Website-1	0.00046	0.00756	0.01039	0.00031	0.00169	0.00143
HTML Website-2	0.02786	0.03097	1.03771	0.00207	0.00197	0.08632

Dari Tabel 2 ditampilkan waktu yang dibutuhkan dalam satuan detik untuk melakukan pencarian kata maupun kalimat pada suatu HTML *website*. Didapatkan perbedaan kinerja yang cukup signifikan diantara kedua algoritma yang telah diuji. Kinerja pencarian menggunakan algoritma BM mencatatkan waktu yang lebih cepat dalam semua kasus pengujian dibandingkan algoritma KMP.

Tampilan teknis *output* program ditunjukkan pada Gambar 4 yang menampilkan waktu hasil pencarian semua kasus yang diuji, termasuk menampilkan keberhasilan algoritma dalam melakukan pencarian dan menampilkan posisi kata kunci yang dicari pada nilai indeksnya, serta kata kunci yang dapat dicari.

Hasil rekapitulasi kinerja masing-masing algoritma dalam pencarian dimasukkan ke dalam file hasil .txt seperti yang ditunjukkan pada Gambar 5. Proses pencocokan dicatat dalam file dan waktu rata-rata yang diperlukan untuk melakukan proses pencarian setiap algoritma.

```

Command Prompt
===== KMP Website-1 =====
Kasus 1: Mencari pattern "Judul Top Secret" <Titel> dalam Website Pertama
Kalimat "top secret" telah ditemukan, berada pada index 6.
Kasus 2: Mencari kalimat UFO experts have long suspected the government "dalam <meta>"
Kalimat "ufo experts have long suspected the government" telah ditemukan, berada pada index 213.
Kasus 3: Mencari kalimat dalam konten "extraterrestrial phenomena"
Kalimat "extraterrestrial phenomena" telah ditemukan, berada pada index 294.
===== BM Website-1 =====
Kasus 1: Mencari pattern "Judul Top Secret" <Titel> dalam Website Pertama
Kalimat "top secret" telah ditemukan, berada pada index 6.
Kasus 2: Mencari kalimat UFO experts have long suspected the government "dalam <meta>"
Kalimat "ufo experts have long suspected the government" telah ditemukan, berada pada index 213.
Kasus 3: Mencari kalimat dalam konten "extraterrestrial phenomena"
Kalimat "extraterrestrial phenomena" telah ditemukan, berada pada index 294.
===== KMP Website-2 =====
Kasus 1: Mencari pattern "Judul video youtube paling trending" <Titel> dalam Website Kedua
Kalimat "video youtube paling trending" telah ditemukan, berada pada index 2105.
Kasus 2: Mencari kalimat paling menarik perhatian masyarakat Indonesia "dalam <meta>"
Kalimat "paling menarik perhatian masyarakat Indonesia" telah ditemukan, berada pada index 2276.
Kasus 3: Mencari kalimat dalam konten "Rahasia Tetangga"
Pencarian tidak menemukan hasil.
===== BM Website-2 =====
Kasus 1: Mencari pattern "Judul video youtube paling trending" <Titel> dalam Website Kedua
Kalimat "video youtube paling trending" telah ditemukan, berada pada index 2105.
Kasus 2: Mencari kalimat paling menarik perhatian masyarakat Indonesia "dalam <meta>"
Kalimat "paling menarik perhatian masyarakat Indonesia" telah ditemukan, berada pada index 2276.
Kasus 3: Mencari kalimat dalam konten "Rahasia Tetangga"
Pencarian tidak menemukan hasil.
    
```

Gambar 4. Output

```

hasil.txt - Notepad
File Edit Format View Help
Kasus: 5
Algoritma yang digunakan: Boyer-Moore
Kalimat pencarian: "ufo experts have long suspected the government"
Ketemu?: true
Waktu pencarian: 5ms
Melakukan matching sebanyak: 59
Rata2 melakukan matching per karakter: 0.001698134929772047

Kasus: 6
Algoritma yang digunakan: Boyer-Moore
Kalimat pencarian: "extraterrestrial phenomena"
Ketemu?: true
Waktu pencarian: 4ms
Melakukan matching sebanyak: 50
Rata2 melakukan matching per karakter: 0.0014390973981119042

===== Data Text HTML Tags#2 =====
Kasus: 1
Algoritma yang digunakan: Knuth-Morris-Pratt
Kalimat pencarian: "video youtube paling trending"
Ketemu?: true
Waktu pencarian: 0ms
Melakukan matching sebanyak: 2149
Rata2 melakukan matching per karakter: 0.02786385737439222

Kasus: 2
Algoritma yang digunakan: Knuth-Morris-Pratt
Kalimat pencarian: "paling menarik perhatian masyarakat Indonesia"
Ketemu?: true
Waktu pencarian: 0ms
Melakukan matching sebanyak: 2389
Rata2 melakukan matching per karakter: 0.030975688816855754
    
```

Gambar 5. Detail algoritma

Dari hasil tes telah diperoleh bahwa kinerja KMP dengan BM memiliki perbedaan kinerja yang cukup signifikan, seperti jika ditemukan hasil pencarian kata kunci untuk situs web Kasus 2-2 misalnya KMP membutuhkan waktu 0,00756 detik dibandingkan dengan kasus yang sama dalam algoritma BM hanya membutuhkan waktu 0,00169 detik dan jika disimpulkan untuk kinerja BM ini adalah 4,47 x lebih cepat dari kinerja KMP.

Adapun skenario terburuk dimana algoritma pencocokan pola ini tidak menemukan kata kunci yang dicari untuk pencarian, waktu yang dibutuhkan algoritma BM adalah 0,08632 detik sementara algoritma harus memakan waktu lebih lama yaitu 1,03771 detik sehingga dapat disimpulkan bahwa algoritma BM, dalam hal ini, lebih cepat 12,02 x daripada algoritma KMP.

#### 4. Kesimpulan

Setelah melakukan serangkaian penelitian dan uji coba pada objek yang ditentukan telah diperoleh, kesimpulan yang diperoleh dari hasil penelitian ini adalah kinerja algoritma BM pada pertama kalinya mengalahkan kinerja algoritma KMP dalam hal ini, terutama dalam skenario "terburuk" dimana kata kunci yang dicari tidak ditemukan. Penelitian lebih lanjut dapat dilakukan dengan basis klien-server dimana akses server file dilakukan oleh klien dari jarak jauh sehingga beban kinerja meningkat dan dapat mempengaruhi kinerja pengujian secara keseluruhan.

#### References

- [1] Bulus H.N., Uzun E., Doruk A., "Comparison of String Matching Algorithms in Web Documents", International Scientific Conference. UNITECH 2017.
- [2] Rahim R., et al., "Visual Approach of Searching Process Using Boyer-Moore Algorithm". J. Phys. Conf.Ser. 930 012001. 2017.
- [3] Rahim R., et al., "A Review: search visualization with Knuth Morris Pratt algorithm". IOP Conf. Ser : Mater. Sci. Eng.237 012026. 2017
- [4] Computer Hope. (2020). Website definition. (Retrieved from computerhope: <https://www.computerhope.com/jargon/w/website.htm>)
- [5] Knuth, Donald; Morris, James H.; Pratt, Vaughan (1977). "Fast pattern matching in strings". *SIAM Journal on Computing*. 6 (2): 323–350. CiteSeerX 10.1.1.93.8147. doi:10.1137/0206024
- [6] Boyer, Robert S.; Moore, J Strother (October 1977). "A Fast String Searching Algorithm". *Comm. ACM*. New York: Association for Computing Machinery. 20 (10): 762–772. doi:10.1145/359842.359859. ISSN 0001-0782
- [7] HTML Living-standard. <https://html.spec.whatwg.org/#is-this-html5?> Accessed 10/07/20.
- [8] Niechai V. 2019. HTML Tags for SEO: to use or not to use?. <https://www.link-assistant.com/news/html-tags-for-seo.html>. Accessed 17/07/20.
- [9] Xia, B. S., & Gong, P. (2015). Review of business intelligence through data analysis. *Benchmarking*, 21(2), 300-311. doi:10.1108/BIJ-08-2012-0050.
- [10] Baker, P. (2000). "Writing a Literature Review". *The Marketing Review*. 1 (2): 219–247. doi:10.1362/1469347002529189
- [11] Meta, S, Brown. "Transforming Unstructured Data into Useful Information", *Big Data, Mining, and Analytics*, Auerbach Publications, pp. 227–246, 2014-03-12, doi:10.1201/b16666-14, ISBN 978-0-429-09529-0, retrieved 2021-05-29
- [12] Cozby, Paul C. (2009). *Methods in behavioral research* (10th ed.). Boston: McGraw-Hill Higher Education.
- [13] Khan, Zia; Bloom, et al. 2009. "A practical algorithm for finding maximal exact matches in large sequence datasets using sparse suffix arrays". *Bioinformatics*. 25 (13): 1609–1616. doi:10.1093/bioinformatics/btp275. PMC 2732316. PMID 19389736.
- [14] Google Webmaster Tools, "Hypertext Markup Language 5", Conference for SEO January 2012.
- [15] *Insights into what the world is searching for -- the new Google Trends*, Insights Search, The official Google Search, September 28, 2012.

