

13_Sindikasi Web menggunakan Teknik Web Scrapping

by Herbertus Yulianton

Submission date: 11-Apr-2023 12:28AM (UTC+0700)

Submission ID: 2060726821

File name: 13_Sindikasi_Web_menggunakan_Teknik_Web_Scrapping.pdf (399.87K)

Word count: 2983

Character count: 18268

Sindikasi Web menggunakan Teknik Web Scrapping untuk Kolaborasi Pemasaran UMKM Handicraft

Sri Mulyani¹, Felix Andreas Sutanto² dan Heribertus Yulianton³

^{1,2,3}Fakultas Teknologi Informasi, Universitas Stikubank Semarang

E-mail:srimulyani@unisbank.ac.id; felix@unisbank.ac.id; heri@unisbank.ac.id

Abstrak

Pemasaran online menjadi salah satu usaha yang dilakukan sebagian besar UMKM pada saat ini. Media web dan sosial media seperti BBM, Facebook, Twitter dan Instagram menjadi pilihan UMKM dalam menyebarluaskan informasi produknya. Namun cara ini juga tidak memecahkan masalah UMKM begitu saja, masalah promosi media internet juga merupakan tantangan tersendiri. Penelitian ini akan mencoba suatu cara promosi lewat web melalui kolaborasi beberapa web yang dimiliki oleh UMKM. Cara yang dimaksud adalah sindikasi web. Sindikasi web pada dasarnya akan mengambil informasi dari suatu web untuk digunakan di web yang lain. Akan diatur mekanisme supaya antar website UMKM dapat saling bertukar informasi dengan mudah. Teknik yang dipakai adalah scrapping web menggunakan Simple HTML DOM Parser. Teknik scrapping lazim digunakan untuk mengambil isi web apabila suatu web tidak menyediakan format RSS (Really Simple Syndication). Simple HTML DOM Parser dapat mengambil isi web dengan cara menemukan format atau tag HTML tertentu yang berisi informasi produk suatu UMKM. Hasil dari penelitian ini adalah suatu program yang dapat digunakan untuk kolaborasi antar web UMKM. Sehingga dapat digunakan untuk UMKM Handicraft dalam memasarkan produknya.

Kata Kunci: Sindikasi Web, HTML Parser, UMKM

PENDAHULUAN

Data pengguna internet di Indonesia sampai tahun 2014 mencapai 88,1 juta. Dari data BPS (Badan Pusat Statistik) pada tahun 2014 jumlah penduduk Indonesia mencapai 252 juta, berarti persentase pengguna internet sudah mencapai 34,9%. Berdasarkan data yang telah disebutkan, internet sangat potensial untuk media perluasan pasar bagi UMKM. Untuk membantu UMKM handicraft dalam meningkatkan pemasaran produknya, salah satu solusi yang dapat dilakukan untuk perluasan pemasaran adalah melalui media web. Web dapat diakses oleh berbagai pengguna tanpa dibatasi ruang dan waktu. Agar penerapannya efektif, diperlukan suatu mekanisme untuk memasarkan web itu sendiri, dengan demikian web akan dikenal dengan lebih cepat.

Sindikasi web adalah suatu bentuk sindikasi yang menyediakan bagian dari suatu situs web untuk digunakan oleh situs lainnya. Biasanya penerapannya digunakan untuk berita. Pendekatan sindikasi web dapat digunakan untuk berbagi data utama pada suatu domain. Sindikasi dapat diterapkan di bidang akademik, komersial, sektor publik dan bidang

yang lainnya. Pada beberapa bidang ilmu telah diberikan suatu standar dalam penyusunan data, contohnya adalah arkeologi, konservasi dan kesehatan. Dalam teknologi web juga telah ada model web service dan semantik web yang memanfaatkan format data yang standar seperti RDF, SPARQL, OWL dan XML. Pendekatan Mashup dapat diterapkan untuk berbagi data tersebut. (Kansa dan Bissell, 2010).

Metode Mashup dapat digunakan untuk mendapatkan suatu berita utama dari sekumpulan berita yang memiliki topik sejenis. Mashup adalah aplikasi yang dapat mengkombinasikan isi dari dua atau lebih sumber berita menjadi satu berita yang baru. Mashup akan mengakses data atau informasi secara langsung dari situs web dan secara programatik menciptakan hasil data yang baru, terintegrasi dan lebih bermanfaat. Komponen mashup ada tiga, yaitu sumber data, algoritma mashup dan platform presentasi. Pembuatan aplikasi di sisi server memiliki empat modul utama untuk mengunduh RSS, mengelompokkan berita, memilih berita utama dan pengelompokan jaringan. Pemilihan berita utama didasarkan pada Google Custom Search.

Google Custom Search adalah platform Google yang dibuat untuk memberikan pelayanan pencarian spesifik kepada pengembang aplikasi web. (Amri dkk, 2012).

Berdasarkan penelitian diatas akan dibangun model pemasaran yang mengikutsertakan partisipasi UMKM Handicraft dalam membentuk komunitas, sedangkan teknik pengambilan informasi akan menggunakan teknik web scrapping dengan Simple HTML DOM Parser. Penggunaan teknik scrapping digunakan agar web yang berinteraksi tidak perlu menyediakan data dengan format khusus seperti XML, RSS atau Atom.

Sindikasi web adalah suatu cara untuk mendapatkan sebagian isi dari suatu web. Cara ini biasanya dilakukan untuk mendapatkan update suatu informasi yang diterbitkan oleh suatu organisasi melalui web. Hal ini dapat dilakukan secara sederhana dengan melisensikan isinya untuk dapat digunakan orang lain. Umumnya, sindikasi web merujuk pada penyediaan umpan web dari suatu situs kepada orang lain untuk memberikan ringkasan isi yang baru ditambahkan pada situs tersebut, misalnya berita atau kiriman forum terbaru. Sindikasi isi web dalam skala besar diawali pada tahun 2001 sewaktu Miniclip secara bebas mensindikasikan permainan interaktif terhubung berbasis penjelajah web kepada umum. Kini berbagai tipe isi telah disindikasikan di Internet. Jutaan penerbit terhubung termasuk surat kabar, situs web komersial, dan blog mempublikasikan tajuk berita terbaru mereka, tawaran produk, atau tulisan blog dalam format standar umpan web.

Sindikasi menguntungkan baik bagi situs yang menyediakan informasi maupun situs yang menayangkannya. Untuk situs penerima, sindikasi isi adalah suatu cara efektif untuk menambah kedalaman dan kemutakhiran informasi pada halaman-halaman mereka dan membuatnya menjadi lebih menarik untuk pengunjung. Untuk situs pengirim, sindikasi memberikan pemaparan di berbagai platform terhubung. Ini akan menghasilkan lalu lintas baru bagi situs pengirim dan membuat sindikasi

menjadi suatu cara pemasaran yang mudah dan gratis.

Semakin diterimanya sindikasi web juga menjadi perhatian bagi pemasar terhubung. Peselancar web menjadi lebih hati-hati dalam memberikan informasi personal untuk bahan pemasaran (seperti mendaftar pada suatu nawala) dan sebaliknya lebih memilih untuk berlangganan pada suatu umpan.

Walaupun format sindikasi dapat berupa apa pun yang dapat dikirim melalui HTTP, seperti HTML atau JavaScript, umumnya sindikasi dikirim dalam format XML. Dua keluarga utama format sindikasi web adalah RSS dan Atom.

RSS singkatan dari Really Simple Syndication (menurut versi terbaru 2.0). Atau sindikasi yang sangat sederhana/simpel, kata sindikasi sendiri menurut bahasa dalam kamus KBBI adalah “gabungan, persekutuan.” Jadi RSS adalah sebuah dokumen file yang digunakan situs web atau web blog untuk himpunan kontennya yang selalu berubah (bertambah). RSS adalah sebuah file berformat XML untuk sindikasi yang telah digunakan (di antaranya dan kebanyakan) situs web berita dan weblog. Singkatan ini biasanya mengarah ke beberapa:

- Rich Site Summary (RSS 0.91)
- RDF Site Summary (RSS 0.9 and 1.0)
- Really Simple Syndication (RSS 2.0)

Teknologi yang dibangun dengan RSS mengizinkan kita untuk berlangganan kepada situs web yang menyediakan umpan web (RSS feed), biasanya situs web yang isinya selalu diganti secara reguler. Untuk memanfaatkan teknologi ini kita membutuhkan layanan Feed. menurut juragancipir.com, RSS Feed ini meliputi posting baru, berita utama, audio baru dan video berformat standar, serta metadata seperti tanggal penerbitan dan nama penulis.

Dalam artikelnya juragancipir.com, menerangkan, Pada situs wordpress, “RSS Feeds beralamat di <http://yoursite.com/feed>. sedangkan pada situs blogger beralamat di <http://yoursite.blogspot.com/feeds/posts/default>

” . Selain itu, ada situs di internet yang berperan sebagai penyimpan umpan semua situs web. Yaitu <http://www.feedburner.com> yang diberdayakan oleh Google. Feed burner akan me-redirect umpan menuju berbagai jenis feed agregator dan feed reader lainnya menjadi sumber umpan. Misalnya, My Yahoo!, Google Reader, News Gator, Google Fusion, My Aol, Bloglines, dan sebagainya“.

Untuk berlangganan RSS Feed tidak diperlukan biaya, Tanpa diperintah, RSS siap mengantarkan informasi relevan dan terbaru yang diinginkan dari dari blog atau situs favorit. Namun, kita biasanya hanya mendapatkan satu baris atau sebuah pengantar dari isi situs berikut alamat terkait untuk membaca isi lengkap artikelnya.

Web scrapping adalah suatu teknik untuk mengekstrak informasi dari suatu website. Teknik ini dapat digunakan untuk mengambil isi sebuah halaman web secara spesifik, misalnya link gambar, isi berita atau sejenisnya. Teknik Scrapping berkaitan erat dengan pengindekan web yang sudah umum dilakukan oleh bot atau web crawler. Tetapi web scrapping lebih berfokus pada transformasi data tidak terstruktur di web, biasanya dalam format HTML, menjadi data terstruktur yang dapat disimpan dan dianalisis dalam database lokal pusat atau spreadsheet. Scraping web juga terkait dengan web otomasi, yang mensimulasikan penjelajahan manusia menggunakan perangkat lunak komputer. Penggunaan web Scraping termasuk perbandingan harga online, pemantauan cuaca, deteksi perubahan situs, penelitian, web mashup dan integrasi data web.

Ada beberapa macam teknik web scrapping, diantaranya adalah Human copy-and-paste, Text grepping and regular expression matching, HTTP programming, HTML parsers, DOM parsing, Web-scrapping software dan sebagainya.

Simple HTML DOM Parser diciptakan oleh S.C Chen. Idenya didapat dari HTML Parser untuk PHP4 yang ditulis oleh Jose Solorzano dan Yousuke Kumakura untuk filter atributnya. Cara kerjanya adalah dengan memanipulasi tag HTML. Untuk

menjalankannya dibutuhkan PHP5 karena ditulis dengan PHP5. Simple HTML DOM Parser akan menemukan suatu tag dalam halaman HTML seperti cara kerja jquery. Parser juga dapat mengekstrak isi dari HTML dalam suatu baris perintah yang mudah. Source code dapat diperoleh di url <http://simplehtmldom.sourceforge.net/>.

Contoh penerapannya adalah sebagai berikut:

```
// Create DOM from URL or file
$html =
file_get_html('http://www.google.com/');
```

```
// Find all images
foreach($html->find('img') as $element)
echo $element->src . '<br>';
```

```
// Find all links
foreach($html->find('a') as $element)
echo $element->href . '<br>';
```

Hasilnya akan memberikan seluruh daftar gambar dan url yang ada pada halaman www.google.com sebagai berikut:

```
/logos/doodles/2016/2016-doodle-fruit-
games-day-4-5657433650233344-hp.gif
/textinputassistant/tia.png
https://www.google.co.id/webhp?tab=ww
http://www.google.co.id/imghp?hl=id&ta
b=wi
http://maps.google.co.id/maps?hl=id&tab
=wl
http://www.youtube.com/?gl=ID&tab=w1
http://news.google.co.id/nwshp?hl=id&ta
b=wn
https://mail.google.com/mail/?tab=wm
https://drive.google.com/?tab=wo
https://www.google.com/calendar?tab=wc
https://www.google.co.id/intl/id/options/
http://translate.google.co.id/?hl=id&tab=
wT
https://books.google.co.id/bkshp?hl=id&t
ab=wp
https://photos.google.com/?tab=wq
https://docs.google.com/document/?usp=d
ocs_alc
https://www.google.co.id/intl/id/options/
```

https://accounts.google.com/ServiceLogin?hl=id&passive=true&continue=http://www.google.co.id/%3Fgws_rd%3Dcr%26ei%3D1BaoV8i6DivUvgTouK-ICA
<http://www.google.co.id/preferences?hl=id>
<http://www.google.co.id/history/optout?hl=id>
/search?site=&ie=UTF-8&q=Olimpiade&oi=dle&ct=2016-doodle-fruit-games-day-4-5657433650233344-hp&hl=id&sa=X&ved=0ahUKEwjx5-KyibHOAhWKL08KHWCDA3MQPQgD/advanced_search?hl=id&authuser=0
/language_tools?hl=id&authuser=0
http://www.google.co.id/setprefs?sig=0_tMuiG4y1eByDnJ6oHyfK4b0rmUg%3D&hl=en&source=home
http://www.google.co.id/setprefs?sig=0_tMuiG4y1eByDnJ6oHyfK4b0rmUg%3D&hl=jw&source=home
http://www.google.co.id/setprefs?sig=0_tMuiG4y1eByDnJ6oHyfK4b0rmUg%3D&hl=ban&source=home
</intl/id/ads/>
<http://www.google.co.id/intl/id/services/>
</intl/id/about.html>
<http://www.google.co.id/setprefdomain?prefdom=US&sig=rEzcE5P6YB1G9fN1w58muyX9ISM%3D>

Pada penelitian ini akan dibangun cara kolaborasi pemasaran antar UMKM Handicraft. Metode sindikasi web diusulkan untuk digunakan dalam kolaborasi web, dan secara otomatis juga membentuk suatu jaringan usaha di dunia maya. Prinsip dasarnya, jika ada yang mengunjungi suatu web maka secara otomatis akan mendapatkan informasi juga tentang web yang lainnya. Semakin banyak kolaborasi, semakin memperbesar peluang UMKM handicraft untuk meningkatkan pemasaran produknya.

METODE PENELITIAN

Dalam merancang model sistem sindikasi antar web UMKM, dan aplikasi sindikasi web menggunakan teknik scrapping yang termudah untuk berbagi informasi antar web UMKM Handicraft. Kemudian membangun aplikasi dengan teknik web scrapping agar dapat digunakan untuk kolaborasi pemasaran membantu meningkatkan pemasaran UMKM Handicraft melalui mekanisme web. Melalui cara ini UMKM yang baru memulai usaha lewat web akan terbantu dengan UMKM yang telah memulai usaha lebih dahulu. Pemasaran produk dapat diperluas karena pengunjung suatu web UMKM akan mendapatkan informasi juga tentang produk UMKM yang lainnya.

1. Teknik Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan cara sebagai berikut:

a. Observasi

Observasi dilakukan di UMKM Handicraft Semarang untuk mendapatkan informasi.

b. Studi Kepustakaan

Studi kepustakaan dilakukan untuk memperoleh literatur-literatur yang relevan dengan obyek penelitian.

2. Metode Perancangan Sistem

Dalam penelitian ini, metode yang digunakan untuk perancangan sistem meliputi:

a. Analisis

Pada tahap ini dicari alternatif-alternatif pemecahan masalah yang paling sesuai untuk mengatasi permasalahan yang ada. Melakukan pemilihan modul yang sesuai dengan kebutuhan sistem.

b. Desain

Merancang sistem pemecahan masalah untuk menentukan langkah-langkah operasi, prosedur, sekaligus membuat desain sistem secara menyeluruh yang meliputi database dan antarmuka sistem.

c. Implementasi

Implementasi sistem yang telah dibuat, sesuai dengan spesifikasi yang telah ditentukan dalam desain sistem. Membuat program sesuai dengan desain.

d. Testing

Setelah program selesai dibuat maka tahap selanjutnya adalah menguji coba aplikasi.

HASIL DAN PEMBAHASAN

1. Analisa Simple HTML DOM Parser

Simple HTML DOM Parser pada dasarnya akan mencari struktur HTML tertentu pada suatu halaman web. Penggunaannya dimulai dengan menentukan DOM. Document Object Model (DOM) adalah object model standar untuk HTML dan XML yang bersifat platform independent. Menentukan DOM dapat dilakukan dengan cara mudah, yaitu alamat URL dari website yang akan diambil sebagian isinya. Perintah untuk membuat suatu DOM dari url tertentu dapat dilakukan dengan cara sebagai berikut:

```
$html = file_get_html('url');
```

Fungsi file_get_html() didefinisikan oleh Simple HTML DOM Parser sebagai berikut:

```
function file_get_html(){
    $dom = new simple_html_dom;
    $args = func_get_args();
    $dom->load(call_user_func_array('file_get_contents', $args), true);
    return $dom;
}
```

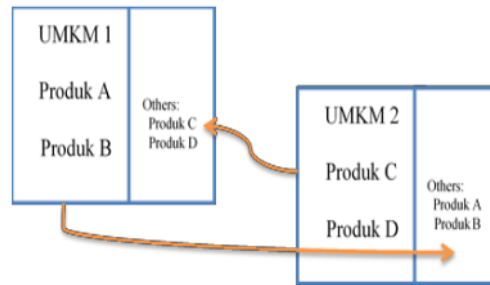
Setelah menentukan DOM, hal yang dilakukan adalah menemukan tag HTML yang akan diambil isinya. Hal ini dilakukan dengan cara menganalisa kode program atau HTML website. Untuk menemukan tag HTML dengan mudah dapat dilakukan seperti contoh sebagai berikut:

```
foreach($html->find('div.article') as $article){
    $item['title'] = $article->find('div.title', 0)->plaintext;
```

```
$item['intro'] = $article->find('div.intro', 0)->plaintext;
$item['details'] = $article->find('div.details', 0)->plaintext;
$articles[] = $item;
}
```

2. Model Sindikasi Web

Penelitian ini dikhususkan untuk UMKM Handicraft. Untuk percobaan telah dipilih dua UMKM yang berbeda jenis produknya, yaitu Boneecha dan Rumah Kayu. Boneecha memiliki produk boneka batik, sedangkan Rumah Kayu memproduksi pernak-pernik dari kayu. Scrapping akan dilakukan dari dua website UMKM tersebut. UMKM Boneecha akan mengambil produk UMKM Rumah Kayu dan sebaliknya. Hasil scrapping akan ditampilkan pada masing-masing halaman web. Untuk memberikan gambaran tentang sindikasi yang dilakukan, dapat dilihat pada gambar 1.



Gambar 1. Model Sindikasi Web

3. Model HTML dan CSS

Agar Simple HTML DOM Parser dapat digunakan dengan mudah, salah satu caranya adalah dengan membuat format HTML yang mudah ditemukan. Saat ini penggunaan tag HTML telah digabungkan dengan CSS agar mendapatkan tampilan yang bagus. Sebagian besar tata letak telah diatur oleh CSS. Untuk itu sebagai penanda dari produk dapat menggunakan CSS. Dalam penelitian ini telah dirancang format sebagai berikut:

```
<div class="product">
<div class="col-md-4">
<a href="produk.php">

```

```

</a>
</div>
<div class="col-md-7">
<h3>Nama Produk</h3>
<h4>Harga</h4>
<p>Keterangan Produk</p>
<a class="btn btn-primary" href="
item.html">Buy</a></div></div>
    
```

Dari format HTML tersebut dapat diketahui bahwa suatu produk berada pada tag <div class="product">. Nama produk ada pada tag <h3>, harga pada tag <h4> dan keterangan pada tag <p>. Untuk gambar dan link tinggal mengikuti tag <a> dan . Kelima tag dapat diekstrak menjadi informasi tentang suatu produk UMKM.

4. Rancangan Database

Untuk rancangan database sidikasi web hanya diperlukan satu tabel saja. Struktur tabel seperti tabel 3.1.

Tabel 3.1. Tabel Sindikasi

Nama Field	Type	Keterang
IDSindikasi	Int(5)	Id
NmUmkm	Varchar(50)	Nama
Produk	Varchar(50)	Nama
Harga	Int(7)	Harga
Keterangan	Text	Keteranga
Link	Text	Link
Gambar	Text	Gambar

5. Implementasi Program

Untuk scrapping dengan Simple HTML DOM Parser digunakan bahasa pemrograman PHP. Caranya adalah menemukan tag <div class="product">, kemudian mengekstrak isi HTML yang ada pada tag tersebut. Potongan program cara scrapping dapat ditunjukkan sebagai berikut:

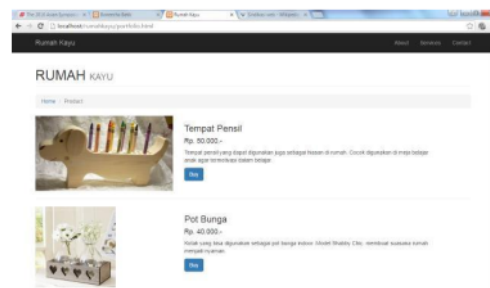
```

<?php
include_once('simple_html_dom.php');
$html =
file_get_html('http://127.0.0.1/rumahkayu'
);
foreach($html->find('div.product') as
$article){
$title = $article->find('h3', 0)->plaintext;
    
```

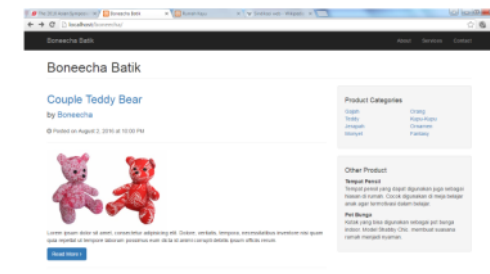
```

$price = $article->find('h4', 0)-
>plaintext;
$details = $article->find('p', 0)-
>plaintext;echo '<p><b>' . $title .
'</b><br>'; echo $details . '</p>';
include_once('simpan.php');
}
?>
    
```

Hasil dari implementasi adalah munculnya produk UMKM lain di website UMKM tertentu. Dalam penelitian ini, produk UMKM Rumah Kayu akan terlihat di UMKM Boneecha. Hasilnya seperti pada gambar 2 dan 3.



Gambar 2. Web Rumah Kayu



Gambar 3. Web Boneecha Batik

KESIMPULAN

Kesimpulan dari penelitian ini adalah:

1. Sindikasi Web dapat digunakan untuk bertukar data antar web UMKM.
2. Penggunaan Simple HTML DOM Parser cukup mudah dilakukan jika kode HTML juga cukup jelas.
3. Kolaborasi pemasaran dapat dilakukan dengan cara menempatkan produk suatu UMKM ke website UMKM lainnya.

SARAN

Saran yang bisa diberikan untuk perkembangan dan penggunaan penelitian ini adalah:

1. Sebaiknya kolaborasi dilakukan pada UMKM yang memiliki produk yang berbeda jenis.
2. Pada penelitian ini belum disusun algoritma untuk filter produk yang sejenis. Kedepannya penelitian dapat dikembangkan untuk mengidentifikasi produk yang mirip, sehingga yang akan ditampilkan hanya produk yang tidak mirip dengan produk UMKM yang menampilkannya.

PUSTAKA

- Amri, Mochamad Gani, Dwi Sunaryono, dan Sarwosri, (2012). Rancang Bangun Aplikasi Pemuat Berita Multi RSS Menggunakan Metode Mashup untuk Mendapatkan Berita Utama dari Kumpulan Berita dengan Topik Sejenis pada Platform Android, JURNAL TEKNIK POMITS Vol. 1, No. 1
- Dwi Agus, (2012). Membangun Model Cyber Cluster Partisipatif Dalam Usaha Memasarkan Secara Online Produk Unggulan UMKM Dan Menarik Investor Di Kabupaten Kendal, Hibah Bersaing DIKTI
- Kansa, Eric C. and Ahrash Bissell, (2010). Web Syndication Approaches For Sharing Primary Data In "Small Science" Domains, Data Science Journal, Volume 9, 8 July 2010.
- Sinta Tridian Galih, dkk., (2011). Rancang Bangun Portal E-Commerce Semarang Handycraft Berorientasi Obyek, Thesis, Magister Sistem Informasi, UNDIP

13_Sindikasi Web menggunakan Teknik Web Scrapping

ORIGINALITY REPORT

24%

SIMILARITY INDEX

24%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

12%

★ media.neliti.com

Internet Source

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On