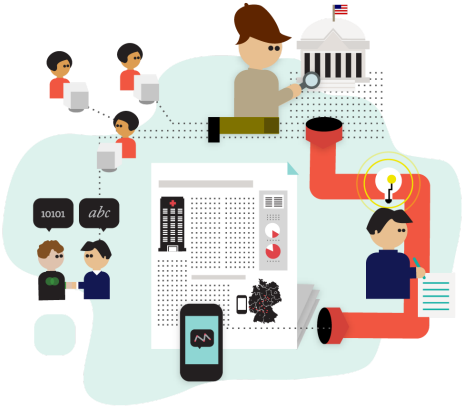
دیتاژورنالیسم

مقدمات و آموزش ابزار جمع‌آوری و تحلیل داده



نسخه اول

مقدمه

این جزوه در راستای آموزش مقدماتی تا متوسط دیتاژورنالیسم برای خبرنگاران و فعالان اجتماعی فارسی زبان تهیه شده است. متن پیش‌رو نسخه پیش‌نویس جزوه می‌باشد که به مرور تکمیل می‌شود. نسخه‌های جدیدتر در اختیار شما قرار خواهد گرفت.

بخش اول
{ تعاریف دیتاژورنالیسم }

کتاب راهنمای کاربردی دیتا ژورنالیسم: روزنامه نگاری به کمک داده ها را دیتا ژورنالیسم می‌نامند. بعضی معتقدند داده بر هر مجموعه‌ای از اعداد اطلاق می‌شود که بیشتر با صفحه گسترده گردآوری می‌شود.

پروفسور نایجل شدبولت (از موسسان data.gov.uk و موسسه داده‌های آزاد): دیتا ژورنالیسم استفاده از داده‌ها و اطلاعات اساسی و عناصری از منابع کلیدی برای شکل دادن یک خبر است.

پاول لویس (سردبیر پروژه‌های خاص، گاردین): دیتاژورنالیسم تنها به معنای وجود داده‌ها یا به دست آوردن و انتشار آن نیست بلکه فرآیندی است که برای تولید آن انجام می‌شود و باید سوالات مناسبی را بپرسید تا پاسخ‌های درست را دریافت کنید.

تعریف کلی

دیتا ژورنالیسم نوعی روزنامه‌نگاری داده محور است که برای انجام آن فرآیندهایی همچون یافتن و گردآوری، پاکسازی و ساماندهی داده‌ها، بیان خبر به یاری یک روایت، تحلیل، تصویرسازی داده‌ها و انتشار داده‌ها طی می‌شود. چرا که.

چند نکته در این تعاریف هست:

1. نگاه فرآیندی است
2. تکیه بر واقعیت‌هاست
3. سوال و رسیدن به جواب مهم است
4. تصویرسازی: درک اطلاعات را برای مخاطب بهتر می‌کند

تاریخچه

در تاریخچه شروع دیتاژورنالیسم بین آمریکا و انگلستان رقابت وجود دارد.

نمونه قدیمی دیتا ژورنالیسم مربوط به سال 1821 و جدولی است که تعداد مدارس، تعداد دانش آموزان و میزان هزینه سالانه تحصیل را نمایش می‌دهد. همچنین در سال 1952 گزارشی به کمک رایانه درباره مرگ و میر در ارتش انگلستان انجام شد و در ادامه از سال1969 رواج استفاده از رایانه برای تحلیل داده‌ها مورد استفاده قرار گرفت. از سال 2000 دیتا ژورنالیسم وارد حوزه مطبوعات شد و به اتاق‌های خبر آمد اما هنوز شکل مدرن نداشت. با آمدن نرم افزارهای جدید در این حوزه از 2009 به بعد دیتاژورنالیسم جدی‌تر شد و غالب مدرن به خودش گرفت؛ در همین سال گاردین به عنوان پیشتاز در این حوزه توسط سیمون راجرز (که الان دبیر داده‌های گوگل است) دیتا بلاگ را راه اندازی کرد که نقش موثری در این حوزه داشت.

دلیل علاقه رسانه‌ها به دیتاژورنالیسم

1. رقابت به ویژه در مقابل شبکه‌های اجتماعی
    1.1. صحت و مقابله با شایعه
    1.2. زمان گذاشتن برای گزارش
    1.3. کشف فساد، تناقض و ....
2. حجم بالای داده‌ها که نیاز به تحلیل برای مخاطب عام دارد

انواع دیتاژورنالیسم

دیتا ژورنالیسم پژوهشی: نگران زمان نیست و زمان زیادی برای کار لازم است؛ حتی ممکن است چند سال طول بکشد، مهارت‌های داده‌ای مثل برنامه‌نویسی باید در حد پیشرفته باشد، به اطلاعات غیر رسمی و محرمانه و داده‌های افشا شده توجه زیادی می‌شود، اصرار بر کنترل مجدد و اکتفا نکردن به یک منبع وجود دارد.

دیتا ژورنالیسم عمومی: بر خلاف دیتا ژورنالیسم پژوهشی در این نوع زمان اهمیت دارد و ممکن است در چند ساعت یا چند روز کار آماده شود، با مهارت‌های داده پایه‌ای مثل excel هم می‌توان کار را پیش برد، برای جمع‌آوری اطلاعات نگاه به مجموعه‌های بزرگ عمومی با دسترسی آسان به داده‌هاست. در رابطه با منابع و کنترل مجدد داده‌ها نسبت به دیتا ژورنالیسم پژوهشی سخت گیری‌کمتری وجود دارد.

نکته: علاوه بر این دو مورد دیتاژورنالیسمی به نام دیتاژورنالیسم همزمان در جریانی ضعیف شکل گرفته که در حال حاضر درحال پررنگ شدن

است. اینجا با کمک الگوریتم‌هایی ربات‌ها به کمک تیم‌های دیتا ژورنالیسم می‌آیند و به تسریع فرآیندها کمک می‌کنند.

فرآیند دیتاژورنالیسم

گردآوری: می‌تواند به دو صورت باشد؛ داده‌ها را داریم و دنبال پرس وجو هستیم تا داده‌ها را به تکامل برسانیم و یا داده و پرسشی وجود ندارد و تلاش می‌کنیم داده جمع آوری شود.

پاکسازی: حذف خطاهای انسانی و تبدیل داده‌ها به ساختاری که با هم هماهنگ شوند.

ایجاد بافت: رفتن در قالب و بستری مناسب؛ مثلا آمار جرایم را داریم ولی به تنهایی معنا ندارد و باید در کنار تعداد جمعیت بیاید تا داده به صورت خالص معنی پیدا کند.

ترکیب: استفاده از دو پایگاه داده است.

انتقال اطلاعات: مرحله آخر که معمولا همراه با تصویرسازی داده است. از مرحله جمع آوری داده تا نهایی شدن کار به ترتیب بر ارزش اطلاعات افزوده می‌شود.

مهارت‌های لازم برای دیتاژورنالیسم

گزارشگری مبتنی بر رایانه: این فرد باید یک ذهن پرسشگر داشته باشد، دنبال داده‌ها برود، برای دستیابی به آنها تلاش کند و حداقلی از آمار و ریاضیات بداند.

توسعه دهنده (دولوپری) اپلیکشن‌های خبری: وظیفه عمومی کردن داده‌ها را دارند.

تصویرسازی داده: وظیفه بصری کردن اطلاعات را برعهده دارند.

نکته: تقریبا 61 نرم‌افزار برای استفاده در حوزه دیتاژورنالیسم وجود دارد که بر اساس کاربردهایی که دارند دسته‌بندی می‌شوند. نرم‌افزارهای تحلیل داده‌ها، نرم افزارهای تصویرسازی داده‌ها و ....

نکته 2: رنگ در دیتاژورنالیسم اهمیت زیادی دارد. به همین دلیل حتی گاهی از نرم‌افزارهای مخصوص کوررنگی که حساسیت بالایی به رنگ دارند استفاده می‌شوند

متخصصین این حوزه اصرار دارند هر رنگی برای دیتاژورنالیسم مناسب نیست و از اصطلاحی به نام رنگ‌های رنگین‌کمانی استفاده می‌کنند که برای نرم افزارها تعریف شده و می‌گویند از آن رنگ‌ها بخاطر تاثیر کمتری که روی مخاطب دارد و مخاطب به آنها عادت کرده است نباید استفاده شود. آنها پیشنهاد می‌کنند پالت‌های رنگی جدید ساخته شود.

مثلا در گاردین برای خودشان جدول رنگی تعریف کردند و تیم دیتاژورنالیسم و افراد فعال در قسمت بصری سازی از این پالت رنگی

استفاده می‌کنند.

ویژگی‌های اعضای تیم دیتاژورنالیسم

1. تیم کوچک باشد. بنا به ضرورت گروه گسترش یابد

2. سخت‌کوشی

3. کلی‌نگری

4. آگاهی از مهارت‌های گوناگون (روابط برای جمع‌آوری خبر، تحلیل روی صفحات گسترده، تصویرسازی و ...)

زمان مورد نظر برای تهیه دیتا ژورنالیسم

در مورد زمان، قانون مطلقی وجود ندارد اما به طور میانگین برای خبر داده‌ای یک ساعت، داده تعاملی یک روز تا یک هفته، اپلیکشن‌ها و ابزار داده‌ای تا دو ماه زمان لازم است.
عوامل موثر بر زمان تهیه: ماهیت پروژه، هدف پروژه، تکنولوژی‌هایی که در کار استفاده می‌شود، عمق تحلیل، ترکیب اعضای تیم و اینکه چه قدر با هم همکاری دارند، سطح تخصص افراد

موضوعات مناسب برای دیتاژورنالیسم
تصور عموم این است که هر آنچه مبتنی بر آمار و داده باشد می‌تواند

مناسب باشد

جذاب‌ترین موضوعات مطرح شده در یک پژوهش: ورزشی، سیاسی (انتخابات)، اجتماعی، مرتبط بودن با دولت، اقتصادی، سرگرم کننده، بهداشت، سلامت و پزشکی.

منابع دیتا: آژانس‌های دولتی، وب‌سایت‌ها، منابع مرتبط با دولت، منابع غیر رسمی، سازمان‌های غیرانتفاعی.

مزایای دیتاژورنالیسم

1. کمک به اینکه یک خبرنگار یا رسانه بتواند ادعای خودش را ثابت کند.

2. امکان پرداختن به موضوعات بزرگ را فراهم می‌کند.

3. می‌تواند مسائل اجتماعی مبهم را با استناد به داده ها روشن کند.

4. اطلاعات ارزشمند را به گفتمان اجتماعی تبدیل کند.

5. موانعی را که سر راه درک درست مخاطب وجود دارد کم کند.

6. داده‌ها معمولا زیبا نیستند و نقش دیتا ژورنالیسم علاوه بر مزیت‌های دیگری که دارد ؛ تبدیل داده‌های زشت به اطلاعات زیباست.

چالش‌های دیتا ژورنالیسم

1. بهره‌گیری نادرست و جهت دار از داده‌ها

2. دانش کم در استفاده از داده‌ها

3. نحوه دستیابی به داده‌ها(در کشورهای مختلف متفاوت است)

4. فرمت داده‌ها

5. هزینه‌ها که بستگی به سطح پروژه دارد

6. خطای انسانی

آثار دیتاژورنالیسم

1. کمک به تأمین منافع مخاطب

2. تأمین منافع رسانه‌ها

3. تأثیر بر دولت،

4. کمک به تعامل دولت و ملت

5. کمک به تعامل میان رسانه و مخاطب

کدام رسانه‌هادر دنیا دارای تیم دیتاژورنالیسم هستند؟

براساس فراوانی بیشتر، گاردین (آثار گاردین الهام بخش و نگاه حاکم در آثارشان انتقادی و افشاگرانه است)، نیویورک تایمز و رسانه‌های انگلستان و آمریکا در رتبه‌های اول هستند.سوئیس، برزیل، آرژانتین، آلمان، فرانسه و ایتالیا در رتبه‌های بعد قرار می‌گیرند. در قاره آمریکا: ایالات متحده، کانادا، برزیل و آمریکای لاتین. در قاره اروپا: انگلستان، آلمان، فرانسه، ایتالیا و کشورهای اسکاندیناوی از کشورهای پیشرو در استفاده از دیتا ژورنالیسم هستند.در کشورهای آسیایی در نپال یک پورتال خبری ویژه دیتا ژورنالیسم راه اندازه شده است. در پاکستان با حمایت‌های بین المللی دوره‌های دیتا ژورنالیسم برگزار می‌کنند. در کشورهای عربی امرالعراقی در این حوزه فعالیت دارد و سایتی هم از سال 2012 راه اندازی کرده است و تلاش می‌کند این مسئله را در کشورهای عربی پررنگ کند. در مجموع دیتا ژورنالیسم در جهان درحال توسعه است، هرچند سرعت رشد متفاوت است.

چند نکته

* در حال حاضر تمایل جدی به دیتا ژورنالیسم در رسانه‌های آنلاین وجود دارد. رسانه‌های دیداری، شنیداری و چاپی هم وارد این حوزه شدند.

* بحث های حقوقی و اخلاقی در دیتا ژورنالیسم مثل دیگر حوزه‌های روزنامه نگاری مطرح است.

* برای گرفتار نشدن در دام اخبار جعلی توصیه می‌شود سواد

رسانه‌ای، سواد بصری و سواد داده مخاطب تا حد امکان افزایش پیدا کند تا در مواجه با این موارد دچار مشکل نشود.

* فرآیندی که از دیتاژورنالیسم می‌توانیم انتظار داشته باشیم در تیم‌های مختلف دنیا متفاوت است اما کلیت ماجرا تفاوتی ندارد. تفاوت در تعریف گام‌ها و اولویت بندی‌هاست.

* همه داده‌ها می‌توانند به توسعه دیتا ژورنالیسم کمک کنند ولی در آینده بیشتر داده‌های بزرگ مورد توجه هستند.

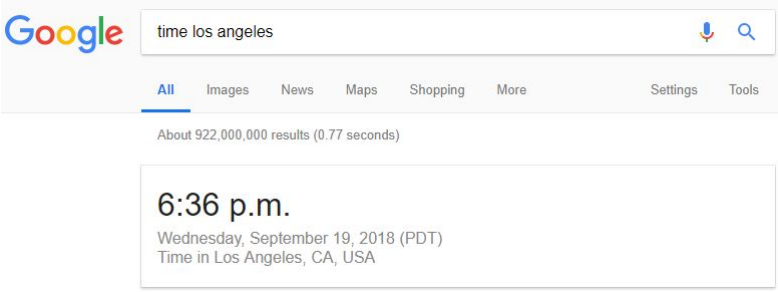
بخش دوم
{ تکنیک‌های جستجو }

امروزه استفاده روزمره از اینترنت بدون وجود جستجوگر گوگل (Google search) غیر ممکن شده است و در انبوه میلیون‌ها صفحه اینترنتی که نمی‌دانید کدام یک دقیقا همان پاسخی است که بدنبال آن می‌گردید جستجوگر گوگل همچون یک معجزه در اختیار شما است تا از سریع ترین راه بتوانید به اطلاعات مورد نظر خود دست پیدا کنید. عمق وابستگی تا آنجاست که حتی دانشجویان برای پیدا کردن ادرس دانشگاه خود در عموم موارد از جستجو در گوگل استفاده می‌کنند.

مشخص کردن ساعت محلی

شاید برای شما هم پیش آمده باشد که بخواهید زمان محلی یک شهر دیگر را بدانید، اما واقعا سرو کله زدن و حساب کردن ساعت محلی از روی ساعت گرینویچ کار خیلی راحتی به نظر نمی رسد. به جای این کار کافی است شما در گوگل نام شهر مورد نظر خود را مشابه زیر وارد نمایید تا ساعت محلی آن به سرعت نمایش داده شود:

time los angeles



همچنین از سایت m¨¨¨¢ÍÙÙ¦©xbYvbÎxf¨ هم می‌توانید برای محاسبه زمان طلوع و غروب خورشید در مکان‌های مختلف استفاده کنید. این اطلاعات در تشخیص زمان حدود تصاویر و فیلم‌ها به شما کمک خواهد کرد.

حذف یک کلمه از نتیجه جستجو

اگر در عبارتی که جستجو می‌کنید، می‌خواهید نتایج مربوط به یک کلمه خاص حذف شود، می‌توانید از روش زیر استفاده کنید:

Terrorism expert -uk -ireland

جستجو برای کلمات هم معنی

اگر می خواهید در نتیجه سرچ خود، کلمات مشابه و هم معنی نیز وارد

شوند باید از علامت ~ مطابق زیر استفاده کنید. در اینجا کلماتی مانند guide , manual , reference نیز در نتیجه جستجو وارد می شوند:

Nodexl installation ~tutorial

معنی کلمات

نیاز نیست برای یافتن معنی یک کلمه یک دیکشنری بر روی کامپیوتر خود نصب کنید و یا زحمت وارد شدن به سایت های دیکشنری را به خود بدهید، با سینتکس (Syntax) زیر می توانید به سادگی معنی کلمات را از طریق گوگل بیابید:

define: yourWord

مثال: define: embezzlement

جستجو برای یک رنج از اطلاعات

اگر به دنبال یک کالا با قیمت های متفاوت می‌گردید می توانید یک رنج از قیمت را به شکل زیر وارد کنید:

Laptop $500..$1000

ماشین حساب

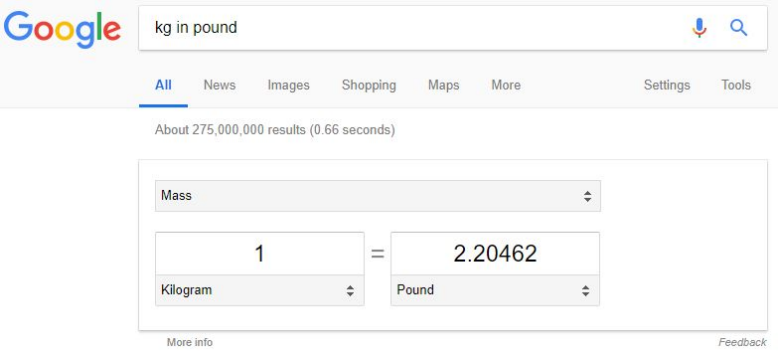
آیا می دانید که جستجوگر گوگل مانند یک ماشین حساب عمل می کند؟ عبارت زیر را در گوگل وارد کنید:

sqrt(10)

تبدیل واحد

اگر می خواهید یک واحد آمریکایی را به یک واحد انگلیسی یا سایر استاندارد های جهانی تبدیل کنید کافیست تا از سینتکسی مطابق شرح زیر استفاده کنید:

kg in pound



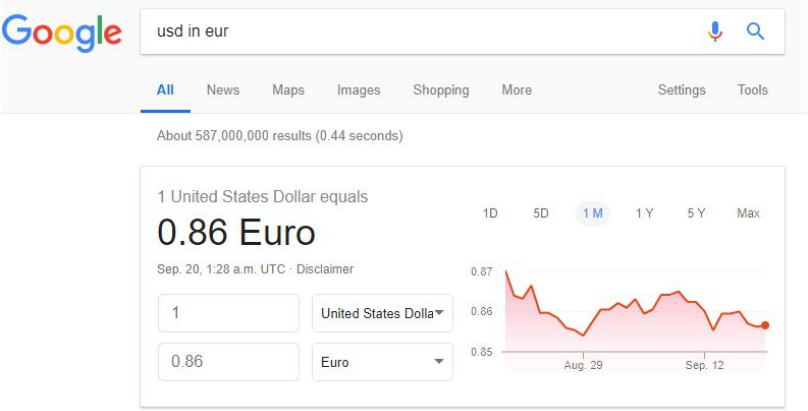
تبدیل نرخ ارز

اگر می خواهید نرخ برابری ارز را مثلا نرخ دلار در برابر یورو ، ریال ، یوان چین یا ارز هر کشور دیگری را بدانید کافی است از عبارتی مانند زیر استفاده کنید:

usd in eur

یا می توانید دقیقا مبلغی را که می خواهید این طور وارد کنید:

100 usd in rmb



جستجو در یک وب سایت خاص از طریق گوگل

خیلی از مواقع استفاده از ابزار جستجو در یک وب سایت چندان کارایی

ندارد و بهتر است از یک موتور جستجوی قوی مثل گوگل برای این امر استفاده کنید، برای این کار کافی است از عبارتی مشابه زیر استفاده کنید:

شهرداری site:asriran.com

جستجوی دقیق یک عبارت

برای اینکه گوگل دقیقا همان عبارتی را که وارد می کنید جستجو کند ، باید عبارت را مانند مثال زیر در میان گیومه قرار دهید:

"مدیرعامل سازمان صدا و سیما"

جستجو بر اساس فرمت فایل

برای آنکه در نتیجه جستجو ، نوع فرمت فایل نیز مد نظر قرار بگیرد می توانید مثال زیر استفاده کنید. به این شکل می توانید تنها فایل های pdf یا پاورپوینت (ppt) و یا سایر فرمت های نوشتاری (document) را جستجو کنید که بسیار در سرعت دستیابی به اطلاعات می تواند مفید باشد.

آمار سلامت روان filetype:pdf

نمایش وضعیت آب و هوا

برای اینکه ببینید در شهر شما و یا یک شهر دیگر وضعیت آب و هوا به چه گونه ای است، کافی است از روش زیر استفاده کنید:

weather los angeles

سایت‌های مرتبط

برای یافتن سایت‌های مرتبط و مشابه یک سایت خاص، می‌توانید از عبارت related استفاده کنید. برای مثال

related:asriran.com

جستجوی تصویر

یکی از ابزارها قوی گوگل، امکان جستجوی تصاویر است (images.google.com). پاسخ این جستجو می‌تواند شامل

موارد زیر باشد:

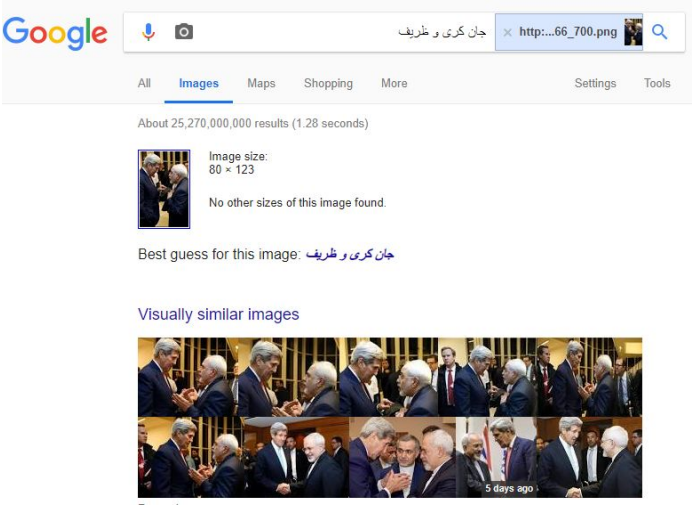
تصاویر مشابه.

سایزهای دیگر تصویر.

سایت‌های شامل این تصویر.

با کلیک روی علامت دوربین می‌توانید از بین گزینه‌های موجود که آپلود از کامپیوتر یا دادن لینک به تصویر روی وب است، یکی را انتخاب کنید.

معمولا گوگل در پاسخ، کلمات تشخیص داده شده با موضوع احتمالی تصویر را نیز به شما نشان خواهد داد که با تغییر آن می‌توانید جستجوی تصویر را دقیق‌تر کنید.



سرویس جستجوی تصاویر جستجوگر روسی یاندکس (https://yandex.com/images/) می‌تواند جایگزین خوبی برای گوگل باشد.

اطلاعات وب‌سایت‌ها

هرچند برای یافتن اطلاعات مالک، سرور و تغییرات وب سایت‌ها از جستجوی گوگل کمک گرفت اما وجود سرویس‌های تخصصی برای این اطلاعات، به شما کمک می‌کند که اطلاعات دقیق‌تر را به صورت دسته‌بندی شده دریافت کنید. برای نمونه می‌توانید از سرویس‌های زیر استفاده کنید:

YÎ m¨¨¢íÙÙ¯mzn¦ÎdzwYnx¨zzv¦ÎbzwÙ

aÎ m¨¨¢íÙÙ®nf¯dx¦ÎnxkzÙ

bÎ m¨¨¢¦ÍÙÙdzwYnxanldY¨YÎbzwÙ

نکته‌ها:

کلمات و حروف اضافه (مانند با، از، و، تا و غیره) را در جستجو استفاده نکنید.

برای جستجو از کلمات کلیدی استفاده کنید و یا کلمات خاصی که حدس می‌زنید در اطلاعات مورد نیازتان وجود دارند.

پیشنهاد می‌شود از ۵ تا ۷ کلمه برای جستجو در گوگل استفاده کنید و بسته به پاسخی که دریافت می‌کنید، کلمات را جایگزین، حذف یا اضافه کنید.

موتورهای جستجوی جایگزین

هرچند موتور جستجوی گوگل کاربران زیادی را به خودش جذب کرده است و با استفاده از زیرساخت گوگل، طیف وسیعی از جستجو را به خوبی پوشش می‌دهد، اما استفاده از موتورهای جستجوی دیگر گاهی می‌تواند پاسخ‌های مناسب‌تری به کاربر بدهد.

- بینگ (https://bing.com):

بینگ یک موتور جستجوی وب متعلق به شرکت مایکروسافت است.

نمایش فهرستی از پیشنهادهای جستجو در زمانی که کاربر واژگان را تایپ می‌کند و همچنین نمایش فهرستی از جستجوهای مرتبط از جمله ویژگی‌های جدید و قابل توجه در آن زمان بود که بینگ ارائه کرد.

- یاندکس (https://yandex.com):

بزرگترین موتور جستجوی روسیه و یکی از بزرگترین موتور جستجوهای جهان و محصول شرکت یاندکس (به روسی:) است.

- داک‌داک‌گو (https://duckduckgo.com/):

در عملکرد خود تا حد زیادی بر داده‌های جمع‌سپاری‌شده همچون ویکی‌پدیا تکیه دارد. این موتور جستجو مدعی است که کاربرانش را ردگیری نمی‌کند. از ویژگی‌های جستجو در این وب‌گاه می‌توان به شیوهٔ جستجوی واژه‌های دارای ابهام در آن اشاره کرد.

بر خلاف برخی از شیوه‌های دیگر جستجو که بر پایهٔ بهترین نتایج در صفحهٔ اول کار می‌کنند، داک‌داک‌گو تلاش می‌کند همهٔ معانی مختلف یک واژهٔ تفسیرپذیر را در صفحهٔ اولش بگنجاند تا نتایج صفحهٔ اول به یک معنی خاص سوگیری نداشته باشد. افزون بر این کاربر را راهنمایی می‌کند تا مقصود خود را بهتر بیان کند.

- ولفرام آلفا (wolframalpha.com):