

فراگستر


اتوماسیون کسب و کار

عنوان سند

قابلیت‌های موتور جستجوی معناگرایانه فراگستر (به‌هجو)


تهران، خیابان مطهری، خیابان کوه نور، خیابان هفتم، پلاک ۱۱
ساختمان فراگستر ☎ ۰۲۱-۴۲۶۲۳۳ 📠 ۰۲۱-۸۸۱۷۲۹۹۸
info@faragostar.net | www.faragostar.net

نسخه پروپوزال: ۱۴۰۱۱۲

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	 اتوماسیون کسب و کار
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		

فهرست مطالب

- ۱- اهمیت جستجو در سازمان و موتور جستجوی سازمانی بهجو ۲
- ۲- ویژگی‌ها و کاربردهای موتور جستجوی معنایی ۳
 - ۱-۲- سهولت در استفاده برای کاربران ۴
 - ۲-۲- سهولت در یکپارچه سازی با منابع اطلاعاتی سازمان ۴
 - ۳-۲- حفظ محرمانگی اطلاعات و رعایت سطوح دسترسی ۵
- ۳- انواع کاربردهای بهجو در سیستم اتوماسیون اداری فراگستر ۶
- ۴- سایر کاربردهای بهجو ۱۰
- ۵- ساختار معماری بهجو ۱۱
- ۶- فناوری های مورد استفاده در بهجو ۱۲

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		

۱- اهمیت جستجو در سازمان و موتور جستجوی سازمانی بهجو

با گسترش استفاده از نرم‌افزارهای اتوماسیون در سازمان‌ها، همه روزه تعداد زیادی اسناد و اطلاعات در سازمان‌های مختلف تولید می‌شود. اسناد تولید شده در سازمان‌ها انواع مختلفی دارد که می‌توان به موارد زیر اشاره کرد:

۱. نامه‌ها و مکاتبات درون سازمانی
۲. نامه‌ها و مکاتبات خارج سازمان
۳. بخش‌نامه‌ها و دستورالعمل‌ها
۴. اسناد فنی و مستندات پروژه‌ها
۵. مستندات مربوط به کارکنان
۶. اطلاعیه‌ها و اخبار درون سازمانی
۷. اسناد و قوانین بالادستی
۸. اطلاعات مشتریان و CRM


با افزایش تعداد و انواع سندهای فوق‌الذکر، جستجوی اطلاعات به یکی از نیازمندی‌های روزمره سازمان‌ها تبدیل شده است. همه‌روزه زمان زیادی از کارمندان به جستجوی اطلاعات موردنیازشان از میان سندها می‌گذرد و چه بسا در بسیاری از موارد نمی‌توانند به سهولت و راحتی سند موردنیاز خود را از میان انبوه سندها بیابند. طبق آمار ارائه شده توسط مکنزی^۱، ۲۰ درصد از زمان کارمندان سازمان، صرف یافتن اطلاعات درون سازمانی می‌شود و در ۳۸ درصد از موارد، جستجوی اطلاعات و اسناد سازمان با شکست روبرو می‌شود.

موتور جستجوی جهانی گوگل، این امکان را میسر کرده است که کاربران بتوانند در میان سندهای وب به جستجو بپردازند و اطلاعات موردنظر خود را بیابند. از سویی دیگر، شرکت‌هایی مانند الکولیا^۲ و موتور جستجوی سازمانی گوگل^۳ (Google Cloud) به شرکت‌های بین‌المللی سرویس جستجوی سازمانی را ارائه می‌کنند. این سرویس‌ها، شباهت زیادی به سرویس گوگل دارند با این تفاوت که به جای وب، اطلاعات درون سازمان را مورد جستجو قرار

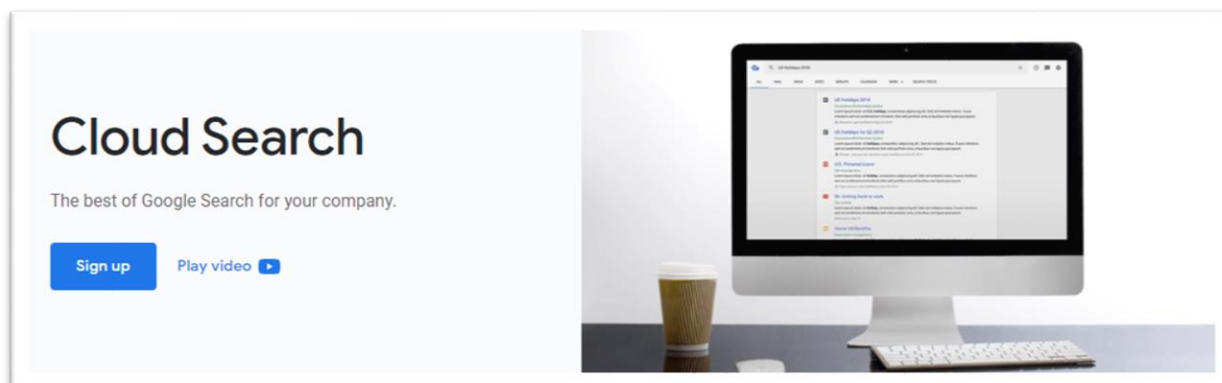
¹ <https://www.mckinsey.com/>

² <https://www.algolia.com/>

³ <https://cloud.google.com/products/search/>

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		


می‌دهند. این سرویس‌ها، عموماً هزینه بسیار بالایی دارند و یا بر روی زیر ساخت ابری ارائه می‌شوند که به دلیل تحریم‌ها و حفظ محرمانگی اطلاعات برای شرکت‌های داخلی قابلیت استفاده ندارند.



شکل ۱- موتور جستجوی سازمانی گوگل که بر بستر ابری ارائه می‌شود

از آنجایی که ابزار مشابهی، هم اکنون در سازمان‌های داخل کشور وجود ندارد که سهولت و کارایی جستجوی اسناد و اطلاعات درون سازمانی را برای کاربران فراهم کند، موتور جستجوی بهجو برای این هدف تولید شده است. بهجو با ارائه‌ی یک واسط کاربری ساده، امکان ارائه خدمت جستجو به کاربران سازمانی را فراهم می‌کند. سادگی در استفاده، یکپارچه سازی سریع با همه منابع اطلاعاتی سازمان، رعایت سطوح دسترسی و امکان یادگیری از رفتار کاربران از مزایای استفاده از بهجو در سازمان‌ها است. بهجو برای تحلیل متون فارسی بهینه‌سازی شده است تا کاربران به کمک این موتور جستجو بتوانند حجم بسیار زیادی از داده‌ها را مورد جستجو و تحلیل قرار دهند. بهجو رتبه‌بندی داده‌ها، اسناد، مدارک و مستندات را به گونه‌ای بهینه می‌کند تا کاربر آنچه به دنبالش می‌گردند را به سادگی پیدا کنند.

۲- ویژگی‌ها و کاربردهای موتور جستجوی معنایی

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		

هدف موتور جستجوی بهجو، ارائه یک واسط کاربری بسیار ساده برای جستجوی اطلاعات سازمانی است. از این رو، این موتور جستجو، ویژگی‌های پایه یک موتور جستجوی وب را ارائه می‌کند که در ادامه به آن‌ها اشاره شده است. بسیاری از این ویژگی‌های بهجو برای کاربران آشنا است، چرا که به صورت روزمره از آنها در موتور جستجو گوگل استفاده می‌کنند. به طور کلی ویژگی‌های موتور جستجو بهجو به سه دسته کلی تقسیم بندی می‌شود.

۱-۲- سهولت در استفاده برای کاربران

موتور جستجوی بهجو با ارائه ویژگی‌های پایه موتورهای جستجو همانند **اصلاح غلط‌های املایی**^۴ (برای مثال کاربر اگر به جای کلمه "ایران"، کلمه "ایراک" را جستجو کند بهجو به صورت خودکار کلمه درست را برای جستجو به کار می‌برد)، **جستجو با کلمات مترادف**^۵ (برای مثال کاربر به جای "نیروی انتظامی جمهوری اسلامی" می‌تواند از ناجا برای جستجو استفاده کند) و **تکمیل خودکار پرس‌وجوی**^۶ (در حین تایپ کاربر، کلمات جستجو به کاربر پیشنهاد داده می‌شود)، تجربه‌ای نزدیک به گوگل برای کاربران ارائه می‌کند.

علاوه بر موارد ذکر شده، **بهینه‌سازی برای زبان فارسی**^۷ یکی از ویژگی‌های مهم این بخش است که به کمک کاربران آمده تا جستجوی اسناد و اطلاعات را ساده و میسر کند. **ریشه‌یابی کلمات** (جستجو با انواع مختلف همزه، الف و ک، جستجوی صورت جمع و منفرد، حذف اعراب در جستجو)، **نرمال‌سازی کلمات** نیز از دیگر ویژگی‌های مهم بهجو برای زبان فارسی است.

برای سازمان‌هایی که دقت جستجو، اولویت بسیار بالایی دارد، با استفاده از رویکردهای یادگیری ماشین (machine learning) این امکان در بهجو وجود دارد که با تعامل کاربران و در طول زمان، کیفیت موتور جستجو افزایش یابد.

۲-۲- سهولت در یکپارچه سازی با منابع اطلاعاتی سازمان

موتور جستجوی بهجو برای مشتریان اتوماسیون اداری فراگستر به صورت آماده^۸ یکپارچه‌سازی شده است. و بدون صرف زمان، قابلیت جستجو بر مکاتبات و اسناد ثبت شده را فراهم کرده است. برای سایر سامانه‌های سازمان هم


⁴ Spell Correction

⁵ Synonym

⁶ Auto Complete

⁷ Persian Language analyser

⁸ Out of the box

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		

امکان یکپارچه‌سازی با استفاده از API های متنوع بهجو امکانپذیر است. از این رو، بهجو می‌تواند همچون یک سامانه استعلام جامع، یک دید ۳۶۰ درجه به کاربران سازمان ارائه نماید.

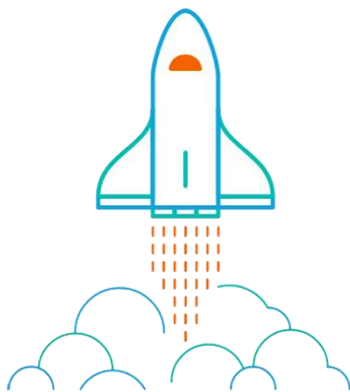
نکته: موتور جستجوی معنایی بهجو نیاز به فعالسازی لایسنس در اتوماسیون اداری فراگستر دارد.

۲-۳- حفظ محرمانگی اطلاعات و رعایت سطوح دسترسی

اگر چه دسترسی سریع و دقیق به اطلاعات سازمان، ارزش زیادی برای کاربران ایجاد می‌کند، به‌همان نسبت نشت اطلاعات نیز می‌تواند عواقب پیچیده‌ای برای سازمان در پی داشته باشد. موتور جستجوی بهجو، با بهره‌گیری از الگوریتم‌های پیچیده، یک توازن بین سرعت دسترسی به اطلاعات و حفظ محرمانگی ایجاد می‌کند. بعبارت دقیقتر در موتور جستجوی سازمانی، هر کاربر پس از ورود به سیستم، به‌سطح تعریف‌شده و مدیریت شده از اطلاعات دسترسی دارد و نمی‌تواند اطلاعات غیرمجاز را رویت نماید.


علاوه بر ویژگی‌هایی که در بالا به آن‌ها اشاره شد، ویژگیهای زیر مزید برتری بهجو نسبت به سایر رقبا را عنوان می‌کند:

- بهجو، خیلی خیلی سریع است



سرعت به سادگی به‌دست نمی‌آید. بهجو برای افزایش سرعت در جستجوی داده‌های متنی از داده‌ساختارهای مناسبی هم‌چون شاخص معکوس دو مرحله‌ای (Inverted Index) بهره می‌برد. از آنجایی که تمام داده‌ها به صورت پیش‌فرض شاخص‌گذاری شده‌اند، کاربر می‌تواند به تمامی داده‌ها با سرعت شگفت‌آوری دسترسی داشته باشد.

- بهجو، مقیاس‌پذیر است

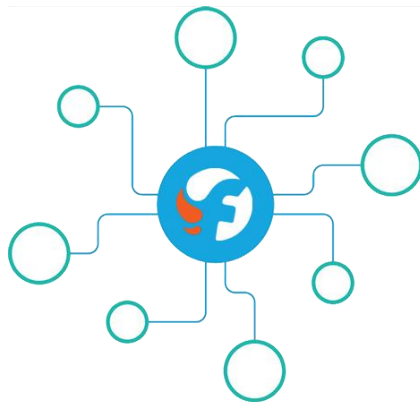
نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		



بهجو به صورت ذاتی مقیاس پذیر است و قابلیت مقیاس پذیری به صورت افقی^۹ را داراست. با افزودن گره‌های جدید به دسته گره‌های فعال، می‌توان با هزینه اندکی این موتور جستجو را برای استفاده در مقیاس‌های بالاتر تجهیز کرد. واسط برنامه نویسی ارتباط با موتور جستجو در مقیاس یک گره تفاوتی با مقیاس صد گره ندارد و تمامی این پردازش‌ها در پشت صحنه انجام می‌گیرد.

در حال حاضر این موتور جستجو جهت استفاده برای بازیابی اطلاعات ذخیره شده در مقیاس اطلاعات یک سازمان ۱۰۰ نفره بر روی یک سرور تا چند Petabytes اطلاعات چند سازمان بر روی چند سرور مقیاس پذیر است.

• بهجو، از کاربران می‌آموزد



در تمام مدت تعامل کاربر با موتور جستجو، اطلاعاتی از قبیل کلمات پرس‌وجوی وارد شده توسط کاربر، کلیک بر روی اسناد و... توسط بهجو ذخیره می‌شود. این داده‌ها هنگامی که به حجم قابل قبولی برسند، توسط الگوریتم‌های یادگیری ماشین برای شخصی‌سازی نتایج جستجوی کاربران استفاده می‌شوند.

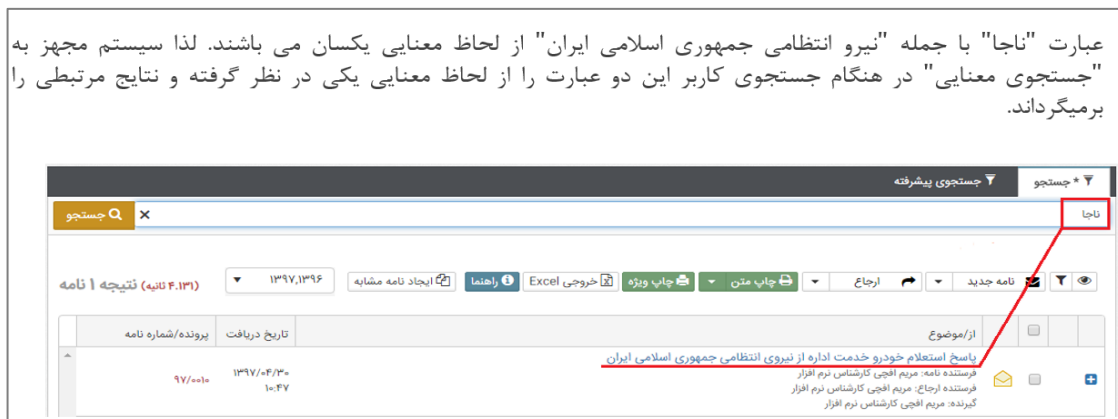
۳- انواع کاربردهای بهجو در سیستم اتوماسیون اداری فراگستر

همانگونه که پیشتر اشاره شد، موتور جستجوی بهجو به صورت پیشفرض با سیستم اتوماسیون اداری فراگستر یکپارچه شده است. در این بخش مثال‌هایی از کاربرد بهجو، در سیستم اتوماسیون اداری فراگستر ارائه شده است.

• بهجو کلمات مترادف را در جستجو در نظر می‌گیرد (Synonym)

⁹ horizontal scalability

ممکن است کاربر در هنگام جستجو، عباراتی که از لحاظ معنایی با عبارت موجود در اسناد برابری می‌کند را وارد نماید. در چنین شرایطی انتظار می‌رود جستجو با همان دقت اولیه انجام شود. موتور جستجوی بهجو این قابلیت را داراست که با دریافت کلمات مترادف در هنگام شاخص‌گذاری اسناد، در هنگام جستجو این کلمات را هم معنی در نظر گرفته و نتایج بهتری را نمایش دهد. بعنوان مثال عبارت "ناجا" می‌تواند به صورت پیش فرض معادل عبارت "نیروی انتظامی جمهوری اسلامی ایران" تعریف شود.



شکل ۲- جستجوی مترادف

• بهجو غلط‌های املائی را اصلاح می‌کند (Spell Correction)

تایپ کردن سخت است و همواره احتمال این که متون تایپ شده (نامه‌ها، مکاتبات، اسم افراد و ...) دارای غلط‌های تایپی و املائی باشند وجود دارد. بهجو به صورت هوشمند توانایی تشخیص و اصلاح غلط‌های املائی کاربران را داراست و هنگامیکه پرس‌وجوی کاربر دارای غلط املائی باشد، پیشنهاد‌های خود جهت تصحیح اشتباه را به کاربر ارائه می‌نماید.



شکل ۳- اصلاح غلط‌های املائی

- بهجو نتایج جستجو را بهینه رتبه‌بندی می‌کند. (Ranking)

در بهجو بر اساس نیازمندی کسب و کار نتایج جستجو می‌تواند رتبه‌بندی (Ranking) قابل تعریفی داشته باشد. بهجو پارامترهای متنوعی از مستندات را برای رتبه‌بندی در نظر می‌گیرد تا همواره مستنداتی که مورد نیاز کاربر است در رتبه‌های اول قرار گیرند. شبیه به گوگل که در اکثر مواقع نیاز شما در همان صفحه نخست پاسخ داده می‌شود.



شکل ۴- ارائه رتبه‌بندی بهینه

- بهجو زبان فارسی را به خوبی می‌فهمد (Normalization)

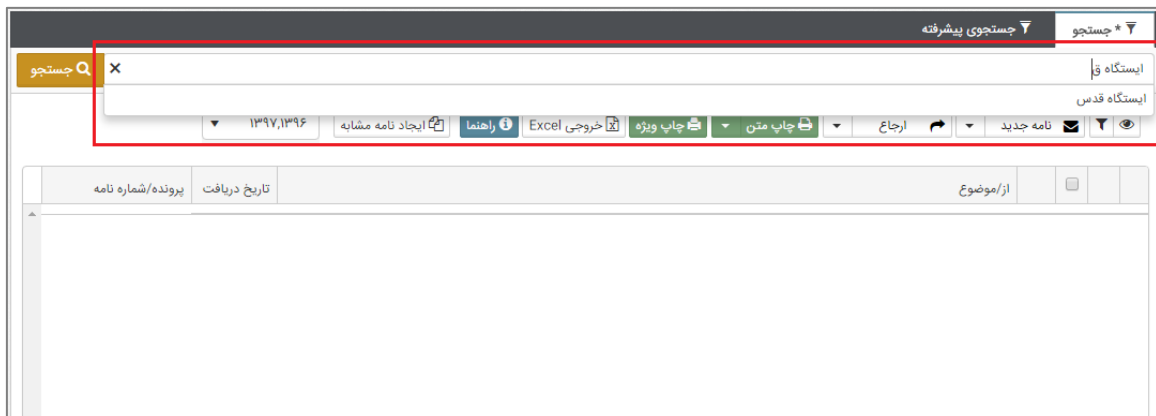
بخش زیادی از چالش‌ها و مشکلات جستجو وابسته به زبان است و برای حل آن‌ها باید از روش‌های پردازش زبان‌های طبیعی بهره برد. بهجو برای زبان فارسی بهینه‌سازی شده است و می‌تواند کلماتی را که به فرم‌های مختلف نوشته می‌شوند اما معنای یکسانی دارند را شناسایی کند. بخشی از کارهایی که در حوزه رفع مشکلات زبان فارسی در بهجو صورت می‌گیرند به شرح زیر است:

- حل مشکل نیم‌فاصله
- حذف اعراب از کلمات
- حذف کشیده از کلمات
- تبدیل حروف کاف و یای عربی به معادل فارسی آن‌ها
- تبدیل اشکال مختلف همزه و الف به یک فرم واحد



شکل ۵- بهینه سازی زبان فارسی

- بهجو پرس و جوی کاربر را به صورت خودکار اصلاح می‌کند (Auto complete)
- در این موتور جستجو هنگامی که کاربر اقدام به جستجو می‌کند، همزمان با تایپ کلمات عبارات مشابهی با عبارت ورودی کاربر از طرف سیستم پیشنهاد می‌گردد که به راحتی کاربر در جستجو کمک می‌کند.



شکل ۶- کامل کردن پرس و جوی کاربر

- بهجو ریشه کلمات را استخراج می‌کند (Stemming)
- پسوندهای رایج زبان فارسی همچون ها، ی نکره، ان و... نباید تاثیری بر نتایج جستجوی کاربران داشته باشند. بهجو به صورت هوشمند این پسوندها را شناسایی کرده و با حذف آن‌ها ریشه کلمات را استخراج می‌کند.

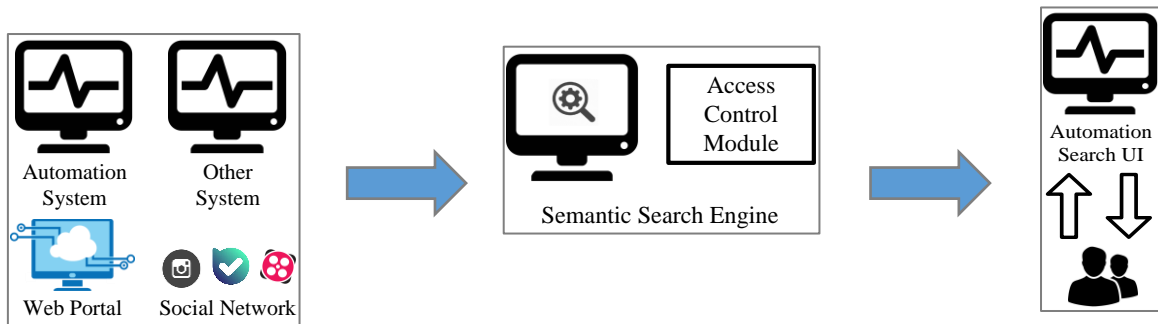


شکل ۷- ریشه یابی کلمات

۴- سایر کاربردهای بهجو

علاوه بر جستجو در سیستم اتوماسیون اداری فراگستر، این امکان وجود دارد که سامانه بهجو به سایر سامانه‌های سازمان متصل شود و جستجوی ۳۶۰ درجه را برای سازمان شما فراهم کند. سامانه‌هایی که قابلیت یکپارچه‌سازی شدن را دارند به شرح زیر است:

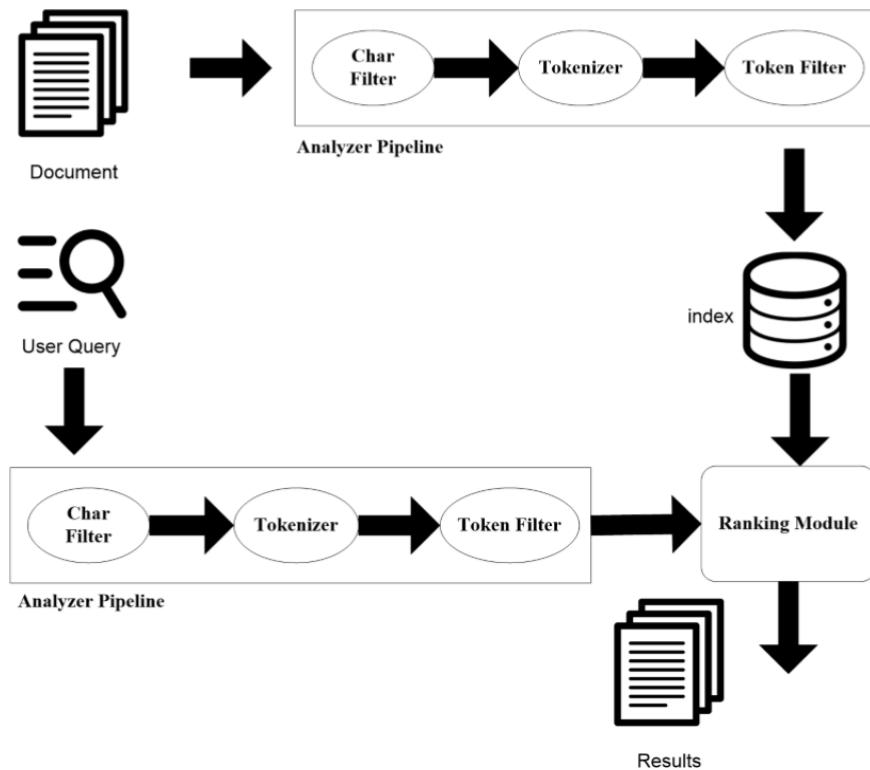
- ۱- سامانه‌های مدیریت تیکت مانند جیرا و ... (جستجوی نام و سوابق تیکت‌ها)
- ۲- سامانه‌های مدیریت روابط مشترکین CRM (جستجوی نام و سوابق مشترکین)
- ۳- سامانه‌های مدیریت نیروی انسانی HR (جستجوی نام و سوابق کارمندان)
- ۴- سامانه‌های تخصصی کسب و کار (جستجو کد ملی، جستجوی کد مشتری و...)
- ۵- سامانه‌های ارشیو الکترونیک (جستجو اسناد فنی سازمانی)
- ۶- سامانه‌ی پرتال داخلی (جستجو اسناد پرتال سازمانی)
- ۷- شبکه‌های اجتماعی سازمان و هشتگ‌های خاص



شکل ۸- سایر کاربردهای بهجو

۵- ساختار معماری بهجو


معماری کلان بهجو و پایپ‌لاین پردازش آن در شکل زیر نمایش داده شده است.



شکل ۹- پایپ لاین پردازش

همان گونه که در شکل بالا مشخص است، جریان سندها، اطلاعات و داده‌ها از سمت بالای عکس (documents) به سمت بهجو انتشار می‌یابد. در پایپ لاین، عملیات مختلفی که هر کدام در یک بیضی نمایش داده شده است به شرح

زیر انجام می‌شود:

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (بهجو)	
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		

۱. Char filter: کاراکترهای فارسی و عربی در این بخش مدیریت می‌شود و بهینه‌سازی‌های مربوط به زبان فارسی در این بخش انجام می‌شود.

۲. Tokenizer: قطعه‌قطعه کردن متن ورودی به کلمات با استفاده از این ماجول انجام می‌شود. مدیریت white-space ها، شناسایی اعداد مهم (مانند شماره تلفن، کد ملی، شناسه مشتری و ...) در این بخش انجام می‌شود.

۳. Token Filter: حذف توکن‌های مزاحم، اصلاح غلط‌های املائی، مدیریت مترادف و ... در این بخش انجام می‌شود. مراحل سه‌گانه بالا، هم بر روی اسناد سازمان اعمال می‌شود و هم بر روی پرس‌وجو کاربران و این امر باعث می‌شود امکان تطبیق پرس‌وجو و سند فراهم شود.


۴. Index: ایندکس چیزی شبیه به جدول پایگاه داده است که اطلاعات مورد نیاز در آنجا ذخیره‌سازی می‌شود تا در زمان مناسب بازیابی آنها انجام شود. در موتور جستجوی بهجو از شاخص‌گذاری دو مرحله‌ای استفاده می‌شود. این فن‌آوری کمک می‌کند، سرعت شاخص‌گذاری افزایش یابد و درعین حال سرعت جستجو نیز بهینه باشد.

۵. Ranking Module: قلب هر موتور جستجو، الگوریتم‌های رتبه‌بندی آن است. ارائه یک رتبه‌بندی بهینه‌کار بسیار پیچیده و دشوار است و تکنیک‌های مختلف یادگیری ماشین و بازیابی اطلاعات در این بخش مورد استفاده قرار می‌گیرد. برای پیاده‌سازی این بخش از کدهای جاوا و R استفاده می‌شود.

۶- فناوری‌های مورد استفاده در بهجو

فناوری‌های مورد استفاده در موتور جستجوی بهجو به بخش‌های زیر استفاده می‌شود:

۱- Core Search Module: برای پیاده‌سازی این بخش از محصول متن باز elastic استفاده شده است.

نسخه سند: ۱۴۰۱۱۲	قابلیت‌های موتور جستجوی معناگرایانه فراگستر (به‌هجو)	 فراگستر اتوماسیون کسب و کار
تاریخ بازبینی: اسفند ماه ۱۴۰۱		
تعداد کل صفحات: ۱۴		

۲- API Integration Module: برای پیاده‌سازی این بخش، برای فراخوانی‌های آسنکرون از RabbitMQ و برای فراخوانی‌های سنکرون از جاوا استفاده شده است.