



وزارت علوم تحقیقات و فناوری

مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

گزارش نهایی طرح پژوهشی

# طراحی و پیاده سازی سامانه ژورنال یاب ایران

مجریان طرح:

دکتر بهاره پهلوان زاده

دکتر محمدباقر دستغیب

## فهرست مطالب

۴	چکیده
۵	۱- ویژگی‌های کلی طرح
۸	۲- مقدمه
۱۲	۳- پیشینه طرح
۱۲	۳-۱- ویژگیهای مشترک الگوریتم های سامانه های توصیه گر
۱۳	۳-۲- طبقه بندی سامانه های توصیه گر
۲۳	۴- معماری و ماژول‌های سامانه توصیه‌گر ژورنال یاب
۲۵	۴-۱- شرح ماژولهای بکار رفته در سامانه
۲۹	۴-۲- معماری سامانه ژورنال یاب
۳۳	۵-آزمون‌ها و نتایج
۳۲	۶- نتیجه‌گیری و پیشنهادات آینده
۳۹	مراجع
۴۱	پیوست ۱- راهنمای استفاده از ژورنال یاب
۴۶	پیوست ۲- رونمایی از سامانه ژورنال یاب ایران

## فهرست اشکال

۹	شکل ۱	واسط کاربری ورود اطلاعات ژورنال یاب
۱۰	شکل ۲	شیوه نمایش ژونال های پیشنهادی
۱۱	شکل ۳	اطلاعات موجود در گزینه بیشتر
۱۸	شکل ۴	صفحه ورود اطلاعات ژورنال یاب الزویر
۱۹	شکل ۵	صفحه نتایج ژورنال یاب الزویر
۲۴	شکل ۶	عملکرد سامانه توصیه گر پیشنهادی
۲۶	شکل ۷	فرآیند نرمال سازی
۲۷	شکل ۸	تحلیل گر آماری
۲۹	شکل ۹	معماری سامانه توصیه گر
۳۲	شکل ۱۰	موتور جستجو
۳۴	شکل ۱۱	نمودار ضریب همبستگی پاسخ سامانه توصیه گر و جواب های واقعی (۴۰ پاسخ برتر)

## چکیده :

این طرح در راستای رسالت و مأموریت مرکز منطقه ای اطلاع رسانی علوم و فناوری و جهت یاری رسانیدن به پژوهشگران، دانشجویان و اساتید کشور ارائه شده است. وجود مجموعه غنی از نشریات فارسی و غیر فارسی در مرکز که همواره در حال ارتقاء و افزایش موجودی است پشتوانه‌ای برای "سامانه ژورنال‌یاب ایران" می‌باشد که با بهره‌گیری از آن‌ها بهترین پیشنهادات را به پژوهشگرا جهت انتخاب ژورنال هدف برای نشر نتایج پژوهشهای خود نماید. این طرح متناسب با اساسنامه و سند راهبردی در راستای هدف استفاده از تکنولوژی ها و شیوه های نوین پردازش برای ارائه برنامه ها و خدمات اطلاع رسانی در منطقه می باشد. همچنین به طور خاص متناسب با اهداف برنامه های شماره ۳ و ۴ و ۵ جدول ۱ "برنامه یک ساله و پنج ساله نظام جامع رایانه ای مرکز" درج شده در فصل هفتم برنامه راهبردی سازمان است.

این سامانه، می‌تواند به عنوان ابزاری جدید جهت کمک به پژوهشگران و دانشجویان استفاده شود. تاکنون مشابه آن توسط برخی از ناشران بزرگ خارجی مانند الزویر، اشپرینگر و ... جهت معرفی ژورنال‌های مناسب برای چاپ مقالات انگلیسی ارائه شده است. چنین سامانه ای برای ژورنال های فارسی نمونه ای نداشته است و پژوهشگران فارسی زبان از این نظر محدودیت داشته اند. "سامانه ژورنال‌یاب ایران" به نویسندگان این امکان را می دهد تا مناسب ترین ژورنال در زمینه موضوعی مقاله تازه نگاشته شده خود را، در مجموعه ژورنال های معتبر ایران بیابند. این سامانه با استفاده از روش‌های بازیابی اطلاعات متنی، عبارات عنوان و چکیده مقاله را در ژورنال های مرتبط با زمینه پژوهشی مقاله جستجو کرده و در نتیجه لیستی از مناسب ترین ژورنال‌ها را برای چاپ مقاله به پژوهشگر پیشنهاد می دهد. در این لیست دید جامعی از کلیه معیارهای ژورنال‌های مرتبط مانند ضریب‌تاثیر، وضعیت رتبه، پایگاه‌های نمایه کننده، وضعیت انتشار آخرین شماره، موضوعات اصلی و فرعی و حوزه موضوعی ژورنال در اختیار محقق قرار می‌گیرد که او را در انتخاب بهترین ژورنال یاری می‌رساند.

## ۱- ویژگی‌های کلی طرح

- ضرورت انجام طرح و اهداف و کاربرد نتایج در رفع تنگناها و مشکلات جامعه و مرکز منطقه

ای:

- سامانه ای برای کمک به پژوهشگران جهت یافتن بهترین ژورنال برای انتشار تحقیقات خود
- پیش بینی میزان شباهت موضوع مقاله پژوهشگران با ژورنال های کشور
- ارائه مشخصات ژورنال ها و ضریب تاثیر آن ها

### • فرضیه ها / سئوالات پژوهش:

- آیا امکان ارائه روشی برای پیش‌بینی میزان شباهت عنوان و چکیده مقاله فارسی با نشریات فارسی کشور (موجود در مرکز منطقه‌ای) وجود دارد؟

- آیا روش ارائه شده برای ژورنال‌یاب لیست نشریات مرتبط با موضوع مقاله را پیدا می‌کند؟

روش پژوهش (جامعه، نمونه، نوع مطالعه، ابزارگردآوری اطلاعات، روش آماری):

مراحل پژوهش به شرح زیر است:

- بررسی روش‌های پردازش زبان طبیعی برای انجام پردازش بر روی متن فارسی
- بررسی روش‌های بازیابی اطلاعات متنی با هدف بازیابی اطلاعات زبان فارسی
- پیاده سازی پردازش‌های متنی مورد نیاز قبل از یافتن میزان شباهت

- پیاده سازی الگوریتم شباهت یاب برای یافتن میزان شباهت عنوان و چکیده فارسی با مجموعه مقالات

موجود در مرکز منطقه‌ای

- پیشنهاد بهترین نشریه از روی شباهت‌های بدست آمده به مقالات

- پیاده‌سازی واسط کاربر تحت وب برای جستجو و نمایش اطلاعات

- روزرسانی اطلاعات کتابشناختی نشریات مرکز منطقه‌ای بر اساس اطلاعات نشریات در سامانه‌های

وزارت علوم و بهداشت

- روزرسانی اطلاعات کتابشناختی نشریات مرکز منطقه‌ای بویژه حوزه موضوعی و ISSN نشریه بر اساس

اطلاعات موجود در سایت نشریات

#### • خروجی‌های این طرح را پس از اجرا مشخص نمایید؟

سامانه‌ای تحت وب می‌باشد که با دریافت عنوان و چکیده مقاله کاربر لیستی از مرتبط ترین نشریات به

موضوع مقاله به کاربر نمایش داده می شود. در این لیست اطلاعات کتابشناختی زیر نیز به تصمیم‌گیری

بهتر کاربر کمک می‌کند :

ضریب تاثیر، وضعیت رتبه، محل اخذ رتبه، موضوع اصلی، موضوع فرعی، دوره انتشار، پایگاه‌ها نمایه-

کننده، حوزه موضوعی، وضعیت انتشار آخرین شماره، کد مصوب کمیسیون وزارت، تاریخ آغاز اعتبار،

صاحب امتیاز، مدیر مسئول، سردبیر

• این طرح پس از اجرا، مورد نیاز کدام سازمان‌ها و... می‌باشد؟

تمام دانشگاه‌های و موسسات تحقیقاتی کشور

## ۲- مقدمه

یافتن نشریه مناسب برای ارسال مقاله یکی از مهمترین مراحل در جریان انتشار مقاله است. برای اکثر نویسندگان این کار به سادگی امکان پذیر نمی باشد، زیرا بسیاری از نشریات دارای موضوعات با تنوع بسیار گسترده ای هستند و بسیاری از مقالات را می توان مقالات بین رشته ای دانست که شامل چندین رشته تخصصی بوده و می توان در نشریات مختلفی در حوزه های مربوطه به چاپ رساند. لذا یافتن نشریات مطابق با موضوع مقاله به عنوان یکی از چالش ها در امر چاپ مقاله مطرح می باشد. به این منظور به سفارش معاونت پژوهش و فناوری وقت مرکز منطقه ای اطلاع رسانی علوم و فناوری (جناب آقای دکتر محمدرضا صالحی) "سامانه ژورنال یاب ایران" در جهت یاری رسانیدن به پژوهشگران، دانشجویان و اساتید کشور پیاده سازی و در مورخ ۱۳۹۶/۹/۶ به آدرس اینترنتی <http://journalfinder.ricest.ac.ir> رونمایی شد. این سامانه، ابزاری جدید جهت کمک به پژوهشگران و دانشجویان می باشد که تاکنون مشابه آن توسط برخی از ناشران بزرگ خارجی مانند الزویر، اشپرینگر جهت معرفی ژورنال های مناسب برای چاپ مقالات انگلیسی ارائه شده است.

"سامانه ژورنال یاب ایران" به نویسندگان این امکان را می دهد تا مناسب ترین ژورنال در زمینه موضوعی مقاله تازه نگاشته شده خود را، در مجموعه ژورنال های معتبر ایران بیابند. بر همین اساس، با عملکردی بسیار ساده جهت جستجوی ژورنال های مناسب، فقط کافیست نویسنده یک مقاله، عنوان و چکیده مقاله خود را در فیلدهای مربوطه وارد نموده تا طی چند ثانیه لیستی از ژورنال های ایرانی مرتبط با مقاله نگارش یافته، معرفی و ارائه گردد.



در شکل های ۱ و ۲ به ترتیب واسط کاربری ورود اطلاعات و شیوه نمایش ژورنال های پیشنهادی نشان داده شده است.

The image shows a web form for submitting a journal article. It is organized into several sections:

- عنوان مقاله** (Article Title): A text input field with a placeholder "عنوان مقاله را اینجا وارد کنید".
- چکیده مقاله** (Article Abstract): A text area with a placeholder "چکیده مقاله را اینجا وارد کنید".
- وضعیت رتبه ژورنال** (Journal Ranking Status): A row of three checkboxes: "علمی پژوهشی" (Scientific Research), "علمی ترویجی" (Scientific Popularization), and "فلقدر درجه علمی" (Academic Rank).
- محل اخذ رتبه ژورنال** (Journal Ranking Location): A row of five checkboxes: "وزارت علوم" (Ministry of Science), "وزارت بهداشت" (Ministry of Health), "دانشگاه آزاد" (Azad University), "حوزه علمیه" (Islamic Studies Center), and "غیره" (Other).
- موضوع مقاله** (Article Subject): A grid of checkboxes for various fields: "علوم کشاورزی" (Agricultural Sciences), "فنی و مهندسی" (Technical and Engineering), "منابع طبیعی" (Natural Resources), "علوم انسانی" (Human Sciences), "پزشکی" (Medicine), "دامپزشکی" (Veterinary Medicine), "علوم پایه" (Basic Sciences), "هنر و معماری" (Art and Architecture), "عمومی" (General), and "غیره" (Other).
- ارسال** (Submit): A blue button with a left-pointing arrow.

شکل ۱- واسط کاربری ورود اطلاعات ژورنال یاب

نام ژورنال	ضریب تاثیر ISC	دوره انتشار	وضعیت رتبه	محل اخذ رتبه	موضوع	نمایه
ادب پژوهی	۰۰۱۴۰	فصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC
ادب فارسی	۰۰۰۷۰	دوفصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC, sid, magiran, ulrich, پایگاه اطلاعاتی علمی دانشگاه تهران
پژوهش زبان و ادبیات فارسی	۰۰۰۷۰	فصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC, SID, Magiran, Sivlca, Noormags
زبان و ادبیات فارسی - دانشکده ادبیات و علوم انسانی دانشگاه تربیت معلم	۰۰۰۵۷	دوفصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC
ادبیات پدیداری	۰۰۰۵۱	دوفصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC, sid, magiran
پژوهش ادبیات معاصر جهان	۰۰۰۲۵	دوفصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC
ادب و زبان	۰۰۰۱۶	دوفصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC, sid, magiran
گرایش نثر و زبان و ادبیات فارسی	۰۰۰۰۰	دوفصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC, SID, Magiran
علوم حدیث	۰۰۰۰۰	فصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC
نثر هنرهای نمایشی و موسیقی	۰۰۰۰۰	دوفصلنامه	علمی پژوهشی	وزارت علوم	هنر و معماری	ISC

شکل ۲- شیوه نمایش ژورنال های پیشنهادی

این سامانه با استفاده از روش های بازیابی اطلاعات متنی و الگوریتم های ترکیبی سامانه های توصیه گر، عبارات عنوان و چکیده مقاله را در ژورنال های مرتبط با زمینه پژوهشی مقاله جستجو کرده و در نتیجه لیستی از مناسب ترین ژورنال ها را برای چاپ مقاله به پژوهشگر پیشنهاد می دهد. در این لیست دید جامعی از کلیه معیارهای ژورنال های مرتبط مانند ضریب تاثیر، دوره انتشار، وضعیت رتبه، محل اخذ رتبه، موضوع و پایگاه های نمایه کننده ژورنال نشان داده شده است. با انتخاب گزینه بیشتر شما می توانید اطلاعات بیشتری از ژورنال را مشاهده کنید. شکل ۳ اطلاعات موجود در گزینه بیشتر را نشان می دهد.



شکل ۳- اطلاعات موجود در گزینه بیشتر

اطلاعاتی مهمی مانند وضعیت انتشار آخرین شماره، موضوعات اصلی و فرعی، حوزه موضوعی ژورنال، صاحب امتیاز و ... به صورت یکجا و متمرکز در اختیار پژوهشگر قرار می‌گیرد و در تصمیم‌گیری نهایی برای محقق بسیار اثرگذار می‌باشد. نتایج مکتسب نشان می‌دهد که سامانه ژورنال یاب می‌تواند از نظر دقت پاسخ‌های ارائه شده و همچنین رتبه‌ی پاسخ‌های مرتبط امتیاز قابل قبولی را کسب نماید.

در ادامه شرح مختصری از ویژگی‌های مشترک الگوریتم سامانه‌های توصیه‌گر ارائه و طبقه بندی سامانه های توصیه‌گر شرح داده می‌شود. پیشینه تحقیقات در بخش سوم و در بخش چهارم معماری سامانه ژورنال‌یاب ارائه شده است. آزمون‌ها و نتایج در بخش پنجم شرح داده می‌شود و در انتها نتیجه گیری بیان می‌گردد.

### ۳- پیشینه طرح :

#### ۳-۱- ویژگیهای مشترک الگوریتم های سامانه های توصیه گر

به طور کلی، در الگوریتم یک سامانه توصیه گر موارد زیر در نظر گرفته می‌شود [1]:

درک محتوا و فیلتر کردن آن: باید از تکنیک هایی برای فیلتر کردن محتوای کم کیفیت از استخر داده های موجود استفاده کرد. توصیه به محتوای کم کیفیت باعث آسیب به سامانه توصیه گر می‌شود. در هر سامانه توصیه‌گر، تعریف کم کیفیت، بستگی به ویژگی آن سامانه دارد.

نمره دهی: بر اساس ویژگی‌های اقلام، باید تابع امتیازدهی برای تخمین احتمال مقادیر و ارزشهای آینده برای نمایش اقلام به کاربر وجود آورد.

رتبه بندی: در نهایت، نیاز به مکانیسمی برای انتخاب یک لیست رتبه بندی شده از اقلام در جهت توصیه نمودن آنها به کاربر است. در ساده ترین سناریو، رتبه بندی ممکن است شامل موارد مرتب سازی بر اساس یک امتیاز واحد برای هر مورد باشد.

به طور خلاصه الگوریتم فوق را می توان به شرح زیر بیان نمود. سیگنال های ورودی بر اساس اطلاعات کاربر و اطلاعات آیتم مورد نظر، توسط مدل های آماری یادگیری ماشین برای تولید نمرات استفاده شده که میزان وابستگی کاربران به اقلام را تعیین می کند. نمره ها توسط ماژول رتبه بندی ترکیب شده تا یک لیست مرتب شده ای از اقلام را بر اساس ترتیب نزولی از اولویت به دست آمده تولید نماید.

### ۳-۲- طبقه بندی سامانه های توصیه گر

در این قسمت به طبقه بندی کلاسیک سامانه های توصیه گر پرداخته می شود:

سامانه های توصیه گر مبتنی بر محتوا<sup>۱</sup>: در این سامانه شباهت اقلام بر اساس ویژگی های مرتبط با موارد مقایسه شده محاسبه می شود. در واقع، فرآیند اصلی انجام شده توسط یک توصیه گر مبتنی بر محتوا، شامل

---

<sup>1</sup> Content-based

تطبیق ویژگی های پروفایل کاربر است که در آن موارد مورد علاقه کاربر با ویژگی های محتوای یک شیء (آیتم)، به منظور توصیه موارد جدید به کاربر ذخیره می شود [2].

فیلترینگ همکاری<sup>۲</sup>: این روش به عنوان ساده ترین و اصلی ترین پیاده سازی های سامانه های توصیه گر محسوب شده که در آن به کاربر جاری و فعال مواردی را که سایر کاربران با سلیقه های مشابه در گذشته علاقه داشته اند را پیشنهاد می دهد [3]. بر خلاف سامانه های توصیه گر مبتنی بر محتوا، فیلترینگ های همکاری می توانند مواردی با محتوای بسیار متفاوت را در صورتیکه پیش از این سایر کاربران علاقه ای به این آیتم های مختلف نشان داده باشند، توصیه کنند [4].

سامانه های محدودیت گر<sup>۳</sup>: این سامانه ها، به عنوان نوع دیگری از سامانه های توصیه گر دانش محور محسوب می شوند. از نظر دانش مورد استفاده، هر دو سامانه مشابه می باشند، تفاوت عمده این دو سامانه در راه حل محاسبه است [5].

---

<sup>2</sup> Collaborative filtering

<sup>3</sup> Constraint-based systems

سامانه های توصیه گر مبتنی بر مورد<sup>۴</sup> : در این سامانه ها که نوع دیگری از سامانه های دانش محور شناخته می شود، توصیه های مبتنی بر معیارهای تشابه مشخص می گردد. در حالی که در سامانه توصیه گر محدودیت گرا عمدتاً از پایگاه های اطلاعاتی از پیش تعریف شده استفاده می شود که حاوی قوانین صریح در مورد نحوه ارتباط نیازهای مشتری با ویژگی های مورد استفاده است [5].

سامانه های توصیه گر ترکیبی<sup>۵</sup> : این سامانه ها بر اساس ترکیبی از تکنیک های ذکر شده فوق ساخته می شود. یک سامانه ترکیبی تلاش می کند که از مزایای یک سامانه برای رفع معایب سامانه دیگر استفاده کند [5].

سامانه های توصیه گر جمعیت شناسی<sup>۶</sup> : در این نوع سامانه، اقلام مبتنی بر جمعیت شناسی توصیه می شود. فرض بر این است که باید توصیه های مختلف برای مدل های جمعیتی مختلف تولید شود. بسیاری از وب سایت ها به راحتی راه حل های شخصی سازی موثر بر اساس جمعیت شناسی را می پذیرند. به عنوان مثال، کاربران به وب سایت های خاصی بر اساس زبان یا کشورشان هدایت می شوند. یا پیشنهادات ممکن است با توجه به سن کاربر سفارشی شود [6].

---

<sup>4</sup> Case-based systems

<sup>5</sup> Hybrid recommender systems

<sup>6</sup> Demographic

سامانه های توصیه گر دانش محور<sup>۷</sup>: در این سامانه ها، اقلام بر اساس دانش حوزه خاص و انطباق ویژگی های آیتم خاص با نیازهای کاربران توصیه می شود. در این سامانه، یک تابع شباهت مقدار نیاز کاربر (شرح مسئله) و تطابق توصیه ها (راه حل های مسئله) را تخمین می زند. در اینجا نمره تشابه می تواند به طور مستقیم به عنوان ابزار توصیه برای کاربر تفسیر شود. سامانه های مبتنی بر دانش در ابتدای پیاده سازی بهتر از سایر سامانه های توصیه گر عمل می کنند، اما اگر این سامانه ها با اجزای یادگیری مجهز نباشند، ممکن است نتایج مطلوبی در اثر گذشت زمان ارائه ندهند [7,8].

سامانه های توصیه گر جامعه گر<sup>۸</sup>: این نوع سامانه های توصیه گر اطلاعات مربوط به روابط اجتماعی کاربران و ترجیحات دوستان کاربر را به دست می آورد. توصیه ها بر اساس رتبه بندی هایی است که توسط دوستان کاربر ارائه شده است. در واقع این سامانه های توصیه گر در پی افزایش شبکه های اجتماعی هستند و امکان دستیابی ساده و جامع از اطلاعات مربوط به روابط اجتماعی کاربران را فراهم می کنند [9]. در ادامه، مروری بر پژوهشهای انجام شده در این حوزه ارائه می گردد:

---

<sup>7</sup> Knowledge-based systems

<sup>8</sup> Community-based



سامانه‌ی نشریاب الزویر<sup>۹</sup> از جمله سامانه‌های شبیه به سامانه ژورنال یاب می‌باشد (شکل ۴ و شکل ۵). در سامانه‌ی نشریاب الزویر از الگوریتم سامانه‌های توصیه‌گر استفاده می‌شود [10]. الگوریتم رتبه بندی توصیه گر نشریه به دو بخش تقسیم می شود. بخش اول مطابق با پرس و جو شده ارسال شده به مقالات موجود در پایگاه داده است. برای این منظور از الگوریتم Okapi BM25 استفاده شده است [11]. Okapi BM25 الگوریتمی است که به طور گسترده ای در زمینه بازیابی اطلاعات مورد استفاده قرار می گیرد. این الگوریتم، اسناد را مطابق با ارتباط آنها با یک پرس و جو، جستجو و رتبه بندی می نماید. به طور معمول، ورودی کیسه ای از کلمات<sup>۱۰</sup> و خروجی مجموعه ای از اسناد با نمرات و رتبه بر اساس کلمات پرس و جو در هر سند، بدون در نظر گرفتن رابطه بین اصطلاحات پرس و جو در یک سند، می باشد. بخش دوم الگوریتم رتبه بندی توصیه گر نشریات، نمرات هر یک از مقالات را به نمرات مجلات منتقل می کند. این مرحله به زیر مراحل زیر تقسیم می شود [10]:

۱. یک میلیون مقاله با بالاترین نمره BM25 را از فهرست مقاله رتبه بندی نگه داشته و نشریه و حوزه های علمی نشریه را که هر مقاله به آن تعلق دارد را ، پیدا شود.

---

<sup>9</sup> Elsevier Journal Finder

<sup>10</sup> Bag-of-word

۲. اگر کاربر نهایی قبلا یک دامنه را انتخاب کرده است، تمام اسنادی را که به این دامنه تعلق ندارند حذف می

شود. اگر کاربر نهایی یک دامنه برای متن ورودی را انتخاب نکند، این مرحله حذف می‌گردد.

۳. محاسبه میانگین نمره BM25 در هر نشریه با استفاده از میانگین نمرات تمام مقالات منتشر شده در همان

نشریه.

The image shows a search form for journals. It contains the following sections:

- Paper title:** A text input field with the placeholder "Enter your paper title here".
- Paper abstract:** A larger text area with the placeholder "Copy and paste your paper abstract here." and a small icon in the bottom right corner.
- Fields of research:** A section titled "Optional: refine your search by selecting up to three research fields" containing a grid of checkboxes for various fields: Agriculture, Economics, Materials Science and Engineering, GeoSciences, Humanities and Arts, Life and Health Sciences, Mathematics, Physics, and Social Sciences. There is also a checkbox for Chemistry.
- Filter:** A checkbox labeled "Limit to journals with Open Access options".
- Button:** A blue button labeled "FIND JOURNAL".

شکل ۴ صفحه ورود اطلاعات ژورنال یاب الزویر

## Search results (10)

Journal title	Sort by Match	Impact Factor	CiteScore	Open Access	Review speed	Acceptance	Production speed		
<b>Journal of Visual Communication and Image Representation</b> <a href="#">More metrics details</a> <a href="#">Scope and information</a>									
2.164	2.8	8 weeks	28 %	1 weeks	Optional	24 Months	\$ 2,400	<a href="#">More info</a>	<a href="#">Submit paper</a>
Match	Impact	CiteScore	Review speed	Acceptance	Open Access	Embargo period	OA Fee + License		
<b>Signal Processing: Image Communication</b> <a href="#">More metrics details</a> <a href="#">Scope and information</a>									
2.244	3.16	6 weeks	16 %	1 weeks	Optional	24 Months	\$ 2,400	<a href="#">More info</a>	<a href="#">Submit paper</a>
Match	Impact	CiteScore	Review speed	Acceptance	Open Access	Embargo period	OA Fee + License		
<b>Image and Vision Computing</b> <a href="#">More metrics details</a> <a href="#">Scope and information</a>									
2.671	3.31	13 weeks	15 %	3 weeks	Optional	24 Months	\$ 3,100	<a href="#">More info</a>	<a href="#">Submit paper</a>
Match	Impact	CiteScore	Review speed	Acceptance	Open Access	Embargo period	OA Fee + License		
<b>Computer Vision and Image Understanding</b> <a href="#">More metrics details</a> <a href="#">Scope and information</a>									
2.498	3.82	7 weeks	18 %	11 weeks	Optional	24 Months	\$ 2,400	<a href="#">More info</a>	<a href="#">Submit paper</a>
Match	Impact	CiteScore	Review speed	Acceptance	Open Access	Embargo period	OA Fee + License		
<b>Pattern Recognition</b> <a href="#">More metrics details</a> <a href="#">Scope and information</a>									
4.582	5.36	9 weeks	20 %	1 weeks	Optional	24 Months	\$ 3,100	<a href="#">More info</a>	<a href="#">Submit paper</a>
Match	Impact	CiteScore	Review speed	Acceptance	Open Access	Embargo period	OA Fee + License		
<b>Neurocomputing</b> <a href="#">More metrics details</a> <a href="#">Scope and information</a>									

شکل ۵ صفحه نتایج ژورنال یاب الزویر

در زمینه سامانه های توصیه گر نشریه، در حال حاضر سامانه هایی وجود دارند که جستجوی مقالات مشابه را انجام می دهند [12]. به عنوان مثال، سامانه پابمد<sup>۱۱</sup> تابعی را برای جستجوی رکوردهای مشابه از رکورد های موجود در پایگاه اطلاعاتی مدلاین<sup>۱۲</sup> ارائه می دهد [13] در سامانه ایتی بلاست<sup>۱۳</sup> جستجو در چکیده مقالات را جهت توصیه نشریه انجام می شود [14]. سامانه ماندلی<sup>۱۴</sup> مقالات مشابه را بر اساس مقالاتی که قبلا منتشر

<sup>11</sup> Pubmed

<sup>12</sup> Medline

<sup>13</sup> eTBLAST

<sup>14</sup> Mendeley

شده اند، جستجو کرده اما مجلات را توصیه نمی کنند [15]. هیچ یک از این سامانه‌های امکاناتی برای نشریات فارسی و زبان فارسی ندارند. از این منظر *سامانه ژورنال یاب/ایران* اولین نمونه عملیات سیستم توصیه‌گر نشریه فارسی برای محققین فارسی زبان می‌باشد.

در وبسایت آمازون<sup>۱۵</sup> از الگوریتم های توصیه گر برای شخصی سازی فروشگاه آنلاین برای هر مشتری استفاده می‌شود [16]. پیشنهادات این فروشگاه آنلاین با توجه به علاقه مشتری تغییر می‌کند. برای مثال عناوین برنامه نویسی به یک مهندس نرم افزار و اسباب بازی های کودک را به یک مادر پیشنهاد داده می‌شود [17].

در آمازون از فیلترینگ مبتنی بر آیتم<sup>۱۶</sup> استفاده می‌شود. بر خلاف فیلترینگ مشارکتی سنتی، محاسبات آنلاین این الگوریتم مستقل از تعداد مشتریان و تعداد اقلام موجود در کاتالوگ محصول است. این الگوریتم اقلام با کیفیت بالا را در زمان واقعی توصیه می‌کند. برای تعیین مشابه ترین آیتم برای یک مورد خاص، الگوریتم با ایجاد مواردی که بیشتر مشترکان تمایل به خرید آن کالا را دارند، یک جدول تشابه- کالا ساخته می‌شود. می‌توان ماتریس محصول-به-محصول<sup>۱۷</sup> را با تکرار از طریق تمام جفت ها و ساخت یک متریک شباهت برای هر جفت ایجاد کرد. با این حال، بسیاری از جفت محصولات دارای کاربران مشترکی نیستند و بنابراین این روش در شرایط

---

<sup>15</sup> Amazon.com

<sup>16</sup> item-based

<sup>17</sup> product-to-product

پردازش زمان و استفاده از حافظه ناکارآمد است. در الگوریتم تکرار پذیر زیر روش بهتری را با محاسبه شباهت بین یک محصول واحد و تمام محصولات مرتبط فراهم می‌کند:

For each item in product catalog, I1

For each customer C who purchased I1

For each item I2 purchased by customer C

Record that a customer purchased I1 and I2

For each item I2

Compute the similarity between I1 and I2

می‌توان شباهت بین دو مورد را با روش‌های مختلف محاسبه کرد. اما یک روش معمول استفاده از اندازه‌گیری کسینوسی است، که در آن هر بردار مربوط به یک آیتم است، نه یک مشتری، و اندازه ابعاد بردار  $M$  مربوط به مشتریانی است که این آیتم را خریداری کرده‌اند. این محاسبات آفلاین از جدول تشابه-کالا بسیار زمانبر بوده و در بدترین حالت  $O(N^2M)$  است. در عمل، این زمان به  $O(NM)$  نزدیک تر است، زیرا اکثر مشتریان خرید بسیار کمی دارند.

نمونه برداری از مشتریانی که عناوین با بهترین فروش را خریداری می‌کنند زمان اجرا را با اندکی کاهش در کیفیت، کمتر می‌کند. با توجه به جدول تشابه-کالا، موارد مشابه هر خرید کاربر توسط این الگوریتم پیدا شده، آنرا رتبه بندی کرده و دسته‌ای از آیتم‌ها را جمع‌آوری نموده و سپس موارد محبوب‌ترین یا مرتبط‌ترین را

توصیه می کند. این محاسبات بسیار سریع بوده و بر اساس تعداد مواردی است که کاربر خرید کرده و یا امتیاز داده است [18]. می و همکاران [19] یک سامانه توصیه گر ویدیویی آنلاین را به نام ویدیوریچ<sup>۱۸</sup> پیشنهاد کردند که ویدیوها را مطابق با ویدئو فعلی بدون نیاز به پروفایل کاربر نمایش می دهد. ارتباط ویدیوها با ویژگی های متنی آنها (برچسب ها، کلمات کلیدی) تعیین می شود. فیلم ها با ویژگی های متنی مرتبط هستند. در پژوهش دیگر، کریشناکومار، سامانه توصیه گر ویدیویی آنلاین به نام ریکو<sup>۱۹</sup> را ایجاد کرد. در سامانه ریکو، پروفایلی برای هر کاربر ساخته شده و داده های مورد علاقه کاربر به طور صریح از وی اخذ شده است. در این سامانه، اطلاعات جمع آوری شده از کاربر را با داده های مشابه جمع آوری شده برای سایر کاربران مقایسه کرده و یک لیست از اقلام توصیه شده را تهیه می کند. از پروفایل کاربری برای مطابقت اقلام با علاقه کاربر استفاده می گردد [20].

---

<sup>18</sup> VideoReach

<sup>19</sup> Recco

#### ۴- معماری و ماژول‌های سامانه توصیه‌گر ژورنال یاب

امروزه با گسترش بکارگیری رایانه، کاربرد سامانه‌های توصیه‌گر توسعه یافته است. همچنین با توجه به حجم و انبوه اطلاعاتی که در پیکره‌ها، بانک‌های اطلاعاتی و سطح وب ذخیره شده است، یافتن مدارک یا صفحات مرتبط و پیشنهاد دادن آن به کاربران می‌بایست هوشمندانه و دقیق باشد. هدف از این پژوهش طراحی سامانه‌ای برای پیشنهاد نشریه به پژوهشگران بر اساس اطلاعات کتابشناختی مقاله می‌باشد.

نشریات دارای حوزه‌های گوناگونی هستند، همچنین برخی از نشریات در حوزه‌های بین رشته‌ای فعالیت می‌کنند. بنابراین یافتن نشریه مناسب برای چاپ مقاله علمی می‌تواند یکی از چالش‌های پژوهشگران پس از نگارش متون علمی باشد. با توجه به اینکه پردازش زبان طبیعی برای متون زبان فارسی همواره با چالش‌هایی روبرو است، لذا سامانه توصیه‌گر باید راه‌حلی برای مشکلات خط فارسی در رایانه نیز لحاظ کند. سامانه توصیه‌گر پیشنهادی از روش ترکیبی تحلیل آماری<sup>۲۰</sup> و مقایسه شباهت براساس شباهت مشارکتی استفاده می‌کند.

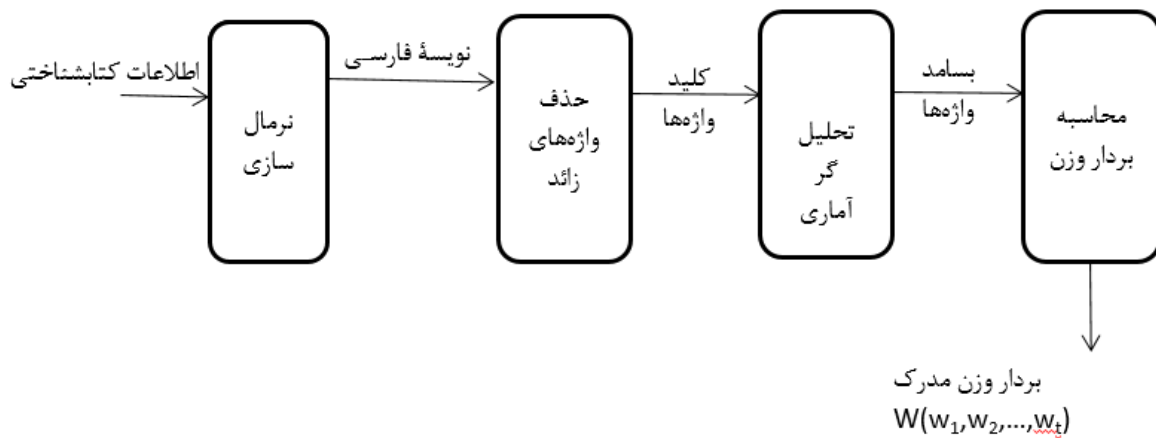
در این روش متن اطلاعات کتابشناختی مقاله مورد نظر کاربر در ابتدا مورد پیش پردازش قرار می‌گیرد. ابتدا نویسه‌های غیرفارسی با معادل‌های آن در زبان فارسی جایگزین می‌شود. سپس واژه‌های زائد<sup>۲۱</sup> از متن حذف

---

<sup>20</sup> Statistical Method

<sup>21</sup> Stop words

می‌شوند. سپس تحلیل‌گر آماری واژه‌های کلیدی مستخرج را از نظر آماری پردازش می‌نماید و برای پرسش بسامد واژه‌های کلیدی را محاسبه می‌کند. بر اساس بسامدهای مستخرج، برداری از وزن واژه‌های کلیدی تشکیل می‌گردد. این بردار نماینده‌ی این اطلاعات کتابشناختی (مدرک) است. شکل ۶ عملکرد این ماژول را به تصویر می‌کشد.



شکل ۶- عملکرد سامانه توصیه گر پیشنهادی

برای محاسبه وزن هر کلیدواژه، از فرمول ۱ برای محاسبه‌ی وزن استفاده می‌شود. همانطور که در فرمول ۱ نشان داده شده است، برای محاسبه‌ی وزن کلید واژه بسامد نرمال شده‌ی واژه در بسامد معکوس مدرک آن واژه ضرب می‌شود تا وزن واژه بدست آید.

$$w_i = tf_i * log_2 df_i \quad (1)$$

برای محاسبه‌ی  $tf_i$  بسامد واژه بر بیشینه بسامد واژه‌ای آن مدرک تقسیم می‌شود. در این صورت مقدار بسامد واژه‌ای به مقداری بین صفر تا یک نرمال می‌شود. بسامد معکوس مدرک نیز با شمارش تعداد مدارکی که شامل



واژه‌ی I هستند محاسبه می‌شود. در این مرحله، هر مدرک با یک بردار مدل می‌شود که نشان دهنده‌ی آن مدرک در فضای برداری است.

پیاده‌سازی مدرک در فضای برداری در مدل پیشنهادی از ترکیب‌های یونی‌گرم<sup>۲۲</sup> و بای‌گرم<sup>۲۳</sup> واژه‌های اطلاعات کتابشناختی مدرک ساخته می‌شود. برای این منظور ترکیب‌های تکی و دو به دو از واژه‌های متن اطلاعات کتابشناختی تهیه می‌شود و بردار نماینده هر مدرک شامل کلیدواژه‌های تکی (یونی‌گرم) و ترکیب‌های دو به دو (بای‌گرم) از واژه‌ها است. روش محاسبه‌ی وزن ترکیب‌های بای‌گرم نیز فرمول (۱) است.

#### ۴-۱- شرح ماژول‌های بکار رفته در سامانه

##### نرمال سازی

پس از دریافت اطلاعات کتابشناختی، لازم است نویسه‌های فارسی نرمال سازی شود. این قسمت برای حفظ دقت سامانه و حذف نویز ناشی از نویسه‌های نادرست بسیار مهم است. در این مرحله ابتدا نویسه‌های عربی به معادل فارسی تبدیل می‌شود. در ضمن جدا نویسی و پیوسته نویسی‌های فارسی با استفاده از قواعد اعمال شده بروی پیکره‌ی واژگان اصلاح می‌شود. برخی از علائم و کاراکترهایی که ممکن است در پکیده و یا عنوان با حروف لاتین و غیر فارسی نمایش داده شده باید و یا فرمول حذف می‌شود تا در انتها متن فارسی یکدست حاصل شود.

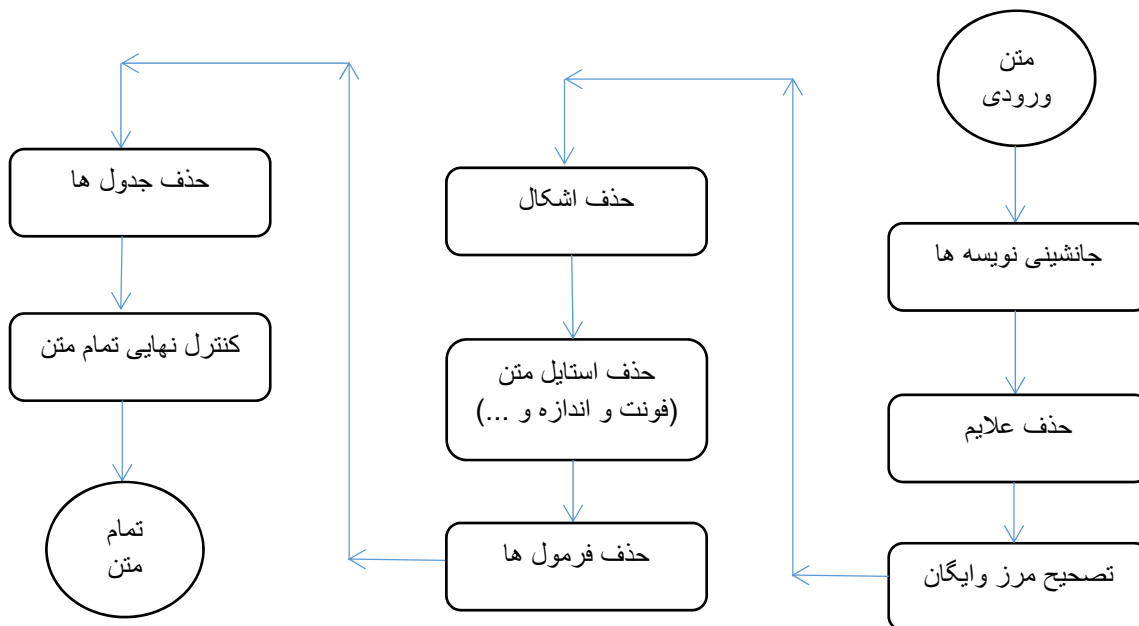
---

<sup>22</sup> Uni-gram

<sup>23</sup> Bi-gram

این مرحله بروی پیکره نیز انجام شده است تا بتوان تطبیق دقیق و جزئی را برای واژگان ورودی و پیکره انجام داد. انجام نرمال سازی فرآیند زمان بری است بنابراین برای تسهیل در جست و جو، این فرایند بروی پیکره بصورت

برون خطی انجام شده است. شکل زیر فرآیند نرمال سازی را مدل می کند.



شکل ۷ فرآیند نرمال سازی

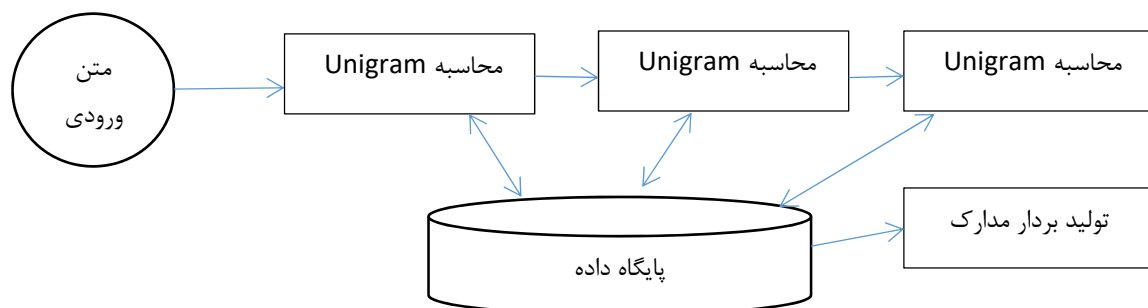
### حذف واژه های زاید

برای محاسبه بردار واژگان که نماینده ی مدرک است، مرسوم است که برخی از واژگان که دارای فرکانس بالایی در پیکره هستند بنابراین خاصیت جداسازی مدارک را ندارند از پیکره حذف می شوند. این کار مزایای زیادی دارد. از جمله مزایای حذف واژگان زائد، کاهش گنجینه لغت (دایره ی واژگان)، کاهش ابعاد بردارها، حذف نویز و افزایش دقت می باشد.

برای این منظور فهرستی از واژگان زائد در فارسی تهیه شده است، سپس با استفاده از یک الگوریتم جست و جوی درون رشته ای ، هر واژه ی زائد با رشته ی تهی (NULL) جانشین می شود. بنابراین پس از این مرحله واژگان زائد از متن حذف و فقط واژگان موثر باقی می ماند.

## تحلیل گر آماری

در این مرحله ابتدا فرکانس تکرار هرواژه در متن (uni-gram) محاسبه می شود. سپس تکرارهای دو واژه در کنار هم (bi-gram) و سه واژه در کنار هم (3-gram) نیز محاسبه می شود. مین عمل بصورت برون خطی برای کل پیکره نیز انجام شده است. با استفاده از مدل زبان و فرکانس های  $TF*IDF$  ، برای هر مدرک برداری از واژگان به طول واژگان گنجینه لغت محاسبه می شود. این بردار نماینده ی هر مدرک است و بنابراین فضای جستجو به فضای برداری تغییر می یابد. در این مرحله با کمک **Sql Server** نیز با استفاده از ترکیب های موجود مدارک شبیه بازیابی می شود.



شکل ۸ تحلیل گر آماری

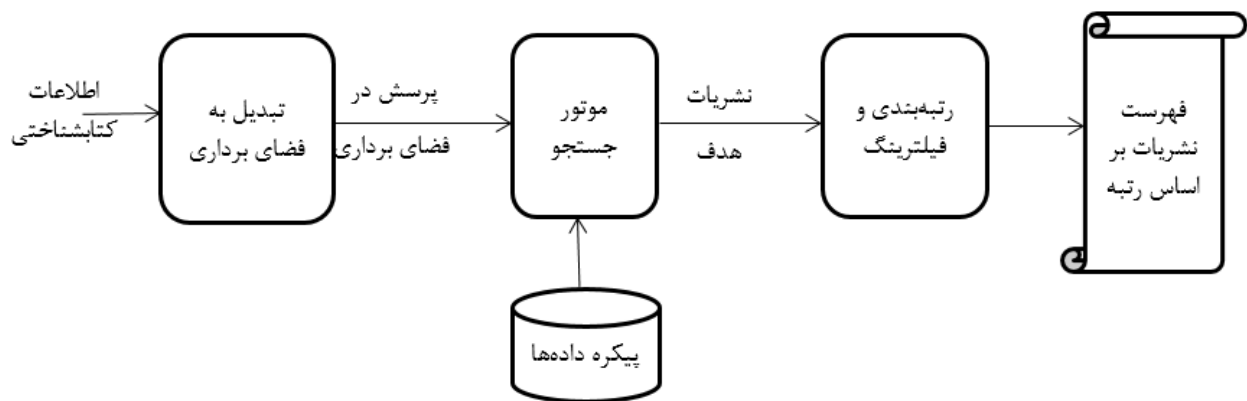
با استفاده از بردارهای آماری محاسبه شده برای هر مدرک یک امتیاز شباهت برای هر مدرک (مقاله) و اطلاعات کتابشناختی ورودی که از کاربر دریافت شده است بدست می آید.

### محاسبه بردار وزن

برای محاسبه بردار وزن ها از روش نرمال سازی استفاده شده است. برای این منظور ابتدا فرکانس واژه های هر مدرک بر اساس ماکزیمم فرکانس واژه های موجود در آن مدرک نرمال می شود. پس از نرمال سازی فرکانس های واژگان مدارک، برای هر واژه در گنجینه لغت، فرکانس مدرک مغلوب IDF نیز محاسبه می شود. سپس با استفاده از این دو معیار و بکار گیری حاصلضرب آنها، وزن هر واژه بدست می آید و نهایتاً بردار هر واژه محاسبه می شود. فضای برداری پیکره بصورت برون خطی و برای اطلاعات کتابشناختی وارد شده بصورت برخط محاسبه می شود.

#### ۴-۲- معماری سامانه ژورنال یاب

پس از محاسبه‌ی بردار مدرکی که به عنوان ورودی داده شده است، باید این بردار با بردارهای مستخرج از پیکره‌ی مقالات مقایسه شود تا بتوان مقالاتی شبیه به این حوزه را شناسایی نمود. اگر مقالات هدف دارای شباهت بالایی به مقاله ورودی باشند، قاعدتاً نشریه‌هایی که این مقالات در آن‌ها چاپ شده‌اند می‌توانند به عنوان نشریات پیشنهادی توسط سامانه‌ی توصیه‌گر پیشنهاد شود. برای این منظور می‌بایست تمام مدارک موجود در پیکره نیز به صورت بردار مدل شوند. بنابراین در یک فرآیند برون‌خطی<sup>۲۴</sup>، از تمامی مدارک پیکره اطلاعات کتابشناختی استخراج می‌شود و سپس با استفاده از ماژول شکل (۶) و محاسبه‌ی ترکیب‌های واژگانی تکی و دو به دو بردار هر مدرک تهیه و ذخیره می‌شود.



شکل ۹- معماری سامانه توصیه‌گر

<sup>24</sup> Off-line

شکل (۹) معماری سامانه‌ی توصیه‌گر را نشان می‌دهد. ابتدا اطلاعات کتابشناختی مقاله مورد نظر کاربر (پرسش) از ورودی دریافت می‌شود. سپس این اطلاعات طبق الگوریتم بکارگرفته در شکل (۶) به برداری از وزن‌ها تبدیل می‌شود. اطلاعات پیکره‌ی مقالات قبلاً و بصورت برون‌خطی به فضای برداری منتقل شده است. این اطلاعات در این مرحله به عنوان بانک مقالات مورد استفاده قرار گرفته و با مقایسه برداری زاویه بین بردار پرسش و مقالات پیکره محاسبه می‌شود. با در نظر گرفتن یک آستانه برای شباهت شبیه‌ترین مقالات به پرسش استخراج شده و از طریق آن‌ها نشریات هدف شناسایی می‌شود. پس از شناسایی نشریات هدف می‌بایست فهرست نشریات براساس پرسش کاربر فیلتر شده و نشریات خارج از دامنه‌ی پرس‌وجو از مجموعه‌ی جواب‌ها حذف شود. در ضمن نشریات باقیمانده نیز بر اساس فاکتورهای چندگانه از قبیل شباهت با مقاله‌ی کاربر، رتبه نشریه در سامانه‌ها رتبه‌بندی و ... رتبه‌بندی می‌شود و در نهایت جواب‌ها بصورت مرتب براساس رتبه‌ی نشریه به کاربر نمایش داده می‌شود.

در سامانه پیشنهادی برای محاسبه رتبه‌ی نهایی از فاکتورهای رتبه مقالات نشریه، رتبه نشریه در نظام رتبه‌بندی و میزان مشابهت موضوعات نشریه با پرسش استفاده می‌شود. برای تهیه معیار واحد، از ترکیب خطی وزن‌دار این سه معیار مطابق با فرمول (۲) استفاده می‌شود.

$$Rank_j = w_1 \cdot R_i^j + w_2 \cdot IF_i^j + w_3 \cdot T_i^j \quad (2)$$

در فرمول (۲)،  $R_i$  نماینده‌ی رتبه کسب شده توسط نشریه با معیار شباهت مقالاتش نسبت به پرسش است،  $IF^j$  نشان دهنده‌ی رتبه‌ی نشریه در نظام رتبه‌بندی نشریات و  $T^j$  نماینده‌ی معیار شباهت موضوعات نشریه با پرسش

کاربر می‌باشد. این ترکیب خطی با استفاده از ضرایب  $W_i$  به صورتی تنظیم می‌شود تا نشریات مرتبط بهترین جواب را بدست آورند. برای این منظور در مجموعه‌ی آموزش ضرایب بصورتی تنظیم می‌شود که جواب نهایی در فهرست سه نشریه برتر بالای لیست مشاهده گردد.

### دریافت اطلاعات کتابشناختی

در این مرحله اطلاعات کتابشناختی دریافت می‌گردد و توسط ماژول نرمال ساز، نرمال سازی می‌شود. در این مرحله واژگان زائد نیز حذف می‌شود و پس از پالایش متن، یک متن فارسی تمیز آماده استفاده توسط موتور جستجو می‌گردد.

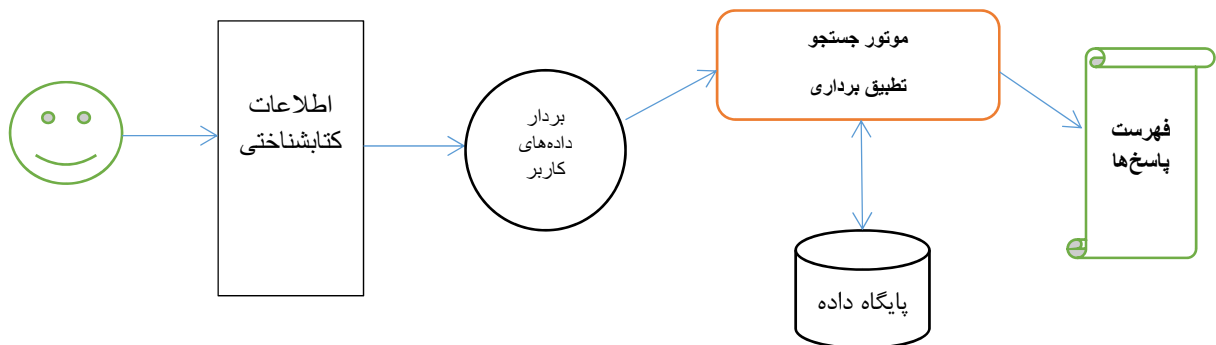
### تولید پیکره داده ها

برای یافتن مجلات هدف باید مقالاتی شبیه به مقاله ی کاربر پیدا و ارزش گذاری شود. برای این منظور پیکره‌ای از اطلاعات کتابشناختی مقالات نشریات فارسی تهیه شده است. این اطلاعات شامل عنوان، چکیده و کلیدواژه‌ها است. متن تمام این داده‌ها توسط واژول نرمال ساز، نرمال سازی شده است و واژگان زائد از آن حذف شده‌اند. پردازش آماری این مقالات نیز بصورت برون خطی انجام شده و نتیجه ذخیره شده است. این پیکره برای جستجوی مقالات شبیه بکار می‌رود. بنابراین لازم است هر مقاله بصورت یک بردار پیاده سازی شود. این برداری تابعی از ویژگی‌های آماری اطلاعات کتابشناختی بوده است.

### موتور جستجو

موتور جستجوی بکار رفته در این سامانه (شکل ۱۰)، با استفاده از بکارگیری فضای برداری و بردارهای تولید شده در مراحل قبل سعی در شناسایی مقالات شبیه به مقاله ی هدف کاربر می‌نماید. برای این منظور با مقایسه بردار مقاله هدف با بردار مقالات پیکره تهیه شده سعی در بازیابی مقالات شبیه می‌نماید. تعداد مقالات با یک آستانه

فیلتر شده و در نهایت نشریات متصل به این مقالات استخراج می شود. اگر فیلتری در جستجوی نشریات هدف توسط کاربر اعمال شده باشد در موتور جستجو اعمال می شود تا مقالات کمتری بازیابی شود. در نهایت یک فهرست از نشریات مرتبط بدست می آید در مرحله ی بعد نشریات ارزش گذاری می شود تا بتوان رتبه ی نشریه را بدست آورد.



شکل ۱۰- موتور جستجو

## رتبه بندی و فیلترینگ

در این مرحله فیلترهای نشریه ای بروی فهرست نشریات بازیابی شده اعمال می گردد. برای هر نشریه نیز یک عدد بین صفر و یک محاسبه می شود که نشان دهنده ی ارزش نشریه در فهرست است. نشریات شبیه تر بر اساس این رتبه بندی در بالای فهرست قرار خواهند گرفت.

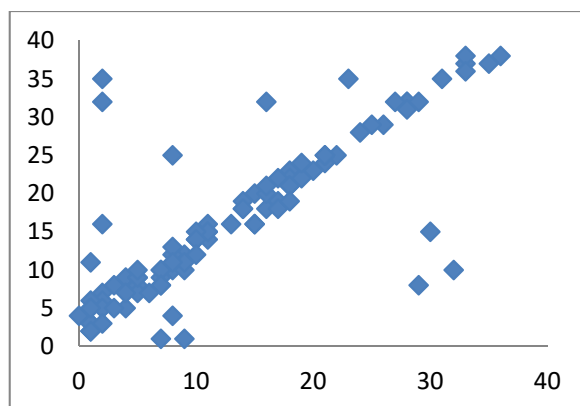
برای رتبه بندی از یک فرمول خطی ترکیبی استفاده می شود. برای محاسبه شباهت هر نشریه میانگین مقالات شبیه در آن نشریه به نشریه هدف در نظر گرفته می شود. بدین ترتیب شباهت هر نشریه محاسبه می شود. رتبه نشریه در سامانه ی ISI و یا ISI و نمایه های بین المللی در صورت وجود و همچنین عدد ارزش نشریه که از مشخصات نشریه استخراج شده است به عنوان سایر معیار ها با معیار شباهت ترکیب شده و یک ترکیب خطی وزن دار را می سازد. در نهایت عدد ارزش هر نشریه با نرمال کردن این فرمول بدست می آید.



## ۵- آزمون‌ها و نتایج

برای آزمون سامانه‌ی توصیه‌گر نیاز است داده‌هایی برای آموزش و آزمایش سامانه گردآوری شود. بنابراین برای این منظور مجموعه داده‌های برچسب‌دار مورد نیاز است که مشخص کند برای هر مقاله نشریه هدف چیست. در این راستا با استفاده از پایگاه‌اطلاعاتی مقالات فارسی مرکز منطقه‌ای اطلاع رسانی علوم و فناوری، مجموعه‌ای دارای حدود ۹ هزار عضو از مقالات علمی پژوهشی که دارای چکیده هستند، انتخاب شده است. برای هر مقاله در این مجموعه، نشریات هدف به صورت خودکار از پایگاه اطلاعاتی استخراج شده‌اند، نشریه‌ای که مقاله در آن چاپ شده است به عنوان مرتبط ترین جواب علامت خورده است.

این مجموعه به دو بخش آموزش و آزمون شکسته می‌شود. بخش آموزش برای تنظیم پارامترهای رتبه‌بند سامانه مورد استفاده قرار گرفته و قسمت آزمون برای آزمایش دقت نهایی سامانه مورد استفاده قرار گرفته است. شکل (۶) نمودار ضریب همبستگی پیرسون را برای داده‌های آزمون و پاسخ‌هایی که سامانه برای مجموعه‌ی آزمون بدست آورده است نشان می‌دهد. همانطور که شکل (۶) نشان می‌دهد، سامانه‌ی توصیه‌گر با همبستگی بالا (۰,۸۱) با جواب‌های واقعی توانسته است پاسخ صحیح را توصیه کند.



شکل ۱۱- نمودار ضریب همبستگی پاسخ سامانه توصیه‌گر و جواب‌های واقعی (۴۰ پاسخ برتر)

جدول (۱) میانگین دقت سامانه‌ی توصیه‌گر را نشان می‌دهد. با مراجعه به داده‌های جدول (۱) مشخص است که سامانه‌ی توصیه‌گر با بکارگیری زیر سامانه رتبه‌بند، می‌تواند پاسخ صحیح را در رتبه‌ی درستی نشان دهد. میانگین رتبه‌ی پاسخ درست برای مجموعه‌ی آزمون ۱۰,۵ است. به عبارت دیگر به طور متوسط در ۱۰ پاسخ اول پاسخ نشریه‌ی مرتبط درست نمایش داده شده است. رتبه‌ی پاسخ مهم است، زیرا اگر پاسخ درست در رتبه‌ی صحیح قرار نگیرد با بالا بردن آستانه فیلترینگ برای نمایش پاسخ، سطح نویز بالا رفته و می‌تواند موجب کاهش دقت سامانه شود. اگر آستانه بسیار سخت‌گیرانه انتخاب شود، ممکن است برخی از توصیه‌های مناسب برای پرسش از فهرست جواب خارج شود. همچنین باید توجه داشت که تعداد جواب‌هایی که می‌توان به کاربر ارائه داد محدود است.

جدول ۱- میانگین دقت سامانه‌ی توصیه‌گر

	۴۰ پاسخ برتر
میانگین دقت (MAP)	۰,۷۹

رتبه‌ی قرارگیری پاسخ در فهرست جواب بسیار مهم است زیرا تعداد پاسخ‌ها براساس آستانه محدود بوده و کاربر انتظار دارد که بالای فهرست جواب‌های دقیق‌تری را مشاهده نماید. در این سامانه‌ی توصیه‌گر کاربر در زمان پرسش دامنه‌ی جست‌وجو را محدود می‌کند (شکل ۴). این محدودیت می‌تواند برای حوزه‌ی نشریه (علوم انسانی، مهندسی، پزشکی و ...)، رتبه‌ی نشریه (دارای ضریب تاثیر، علمی پژوهشی و ...)، بسامد انتشار نشریه و پارامترهای دیگر تنظیم شود. این تنظیمات می‌تواند سامانه‌ی توصیه‌گر را در یافتن جواب صحیح هدایت نماید.

معیار دیگری که برای سنجش صحت فهرست پاسخ سامانه‌ی توصیه‌گر کاربرد دارد، معیار میانگین امتیاز متقابل<sup>۲۵</sup> است (فرمول ۳) [21]. این معیار یک سنجش آماری است که برای هر پردازشی که فهرستی از جواب‌ها را برای یک پرسش تهیه می‌کند، قابل استفاده است.

<sup>25</sup> Mean Reciprocal Rank (MRR)

امتیاز متقابل برای جواب‌های یک پرسش برابر با وارون ضربی رتبه‌ی اولین جواب درست است. میانگین امتیاز متقابل با بدست آوردن میانگین مقدار امتیاز متقابل برای تمام پرسش‌هایی که در مجموعه‌ی  $Q$  قرار دارند بدست می‌آید. در فرمول (۳) مقدار  $rank_i$ ، رتبه‌ی اولین جواب درست در مجموعه‌ی پاسخ‌های پرسش  $Q$  است.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{Rank_i} \quad (۳)$$

با استفاده از این معیار نیز آزمون انجام شده بروی سامانه‌ی توصیه‌گر ارزیابی می‌گردد. و نهایتاً میانگین امتیاز متقابل برای سامانه‌ی توصیه‌گر نشریات (ژورنال یاب) برابر ۰,۵۳ بدست می‌آید.

## ۶- نتیجه‌گیری و پیشنهادات آینده

نتایج بدست آمده نشان می‌دهد که الگوریتم ترکیبی بکارگرفته شده در سامانه‌ی ژورنال یاب کارا است و می‌تواند بخوبی نشریه هدف را شناسایی نماید. همچنین استفاده از ضریب همبستگی پیرسون نشان می‌دهد که پاسخ‌ها با جواب‌های واقعی هم‌راستا بوده و همبستگی خوبی بین پاسخ‌های سامانه و پاسخ‌های واقعی برقرار است.

برای آنکه بتوان از سامانه بصورت عملی استفاده نمود، کاربر باید با محدود سازی دامنه‌ی جستجو، نویز را از مجموعه‌ی پاسخ‌ها حذف نماید که در این صورت دقت پاسخ قابل قبول خواهد بود. دقت پاسخ‌ها برای سامانه‌ی ژورنال‌یاب قابل قبول است. آزمون‌ها نشان داد که دقت بدست آمده در حوزه‌ی نشریات فارسی می‌تواند پاسخ‌های صحیح را در میانگین حدود رتبه‌ی ۱۰ شناسایی نماید. همچنین معیار میانگین امتیاز متقابل نیز نشان می‌دهد که رتبه‌ی پاسخ‌های صحیح در فهرست نتایج، امتیاز قابل قبولی است و میانگین امتیاز کسب شده برای رتبه‌ی پاسخ‌های صحیح قابل قبول است. با توجه به کارآیی سامانه، ژورنال‌یاب می‌تواند برای جستجو در نشریات و یافتن نشریه هدف مورد استفاده پژوهشگران قرار گیرد.

برای توسعه‌ی سامانه در آینده، می‌توان نتایج رتبه‌ی مقالات شبیه به نشریه را با بکارگیری مجموعه منابع مقالات و نشریات و ساخت شبکه معنایی این داده‌ها ارتقاء داد. همچنین با استفاده از مدل مفهومی موضوعی و آنالیز رابطه‌ی پنهانی میان واژه‌ها و نمایه سازی معانی پنهان<sup>۲۶</sup> می‌توان وزن‌دهی بردارها را بهبود بخشید.

موضوع دیگری که می‌توان برای توسعه در آینده در نظر گرفت تعیین دقیق حوزه‌های موضوعی نشریات و بهره‌گیری از آن‌ها برای کمک به پژوهشگر برای محدود کردن بازه نشریات مورد جستجو می‌باشد. با محدود شدن تعداد نشریات مورد جستجو، ژورنال‌یاب پیشنهادات بهتری ارائه می‌دهد. برای بالا بردن دقت حوزه‌های موضوعی نشریات وزارت علوم تحقیقات و فناوری می‌تواند به عنوان متولی، نشریات را به تعیین و رعایت حوزه‌های موضوعی تعیین شده ملزم نماید.

---

<sup>26</sup> Latent Semantic Indexing

- [1] Nichols, D.M., 1997. Implicit ratings and filtering., *Proceedings of the 5<sup>th</sup> DELOS Workshop on Filtering and Collaborative Filtering*, Hungary, pp.31-36.
- [2] Guo, Q. and Agichtein, E., 2012. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior, *proceedings of the 21st international conference on World Wide Web*, pp. 569–578.
- [3] Zhu, Y., He, L., and Wang, X., 2012. User interest modeling and self-adaptive update using relevance feedback technology. *Procedia Engineering*, 29, pp. 721–725.
- [4] Park, Y.J., 2013. An adaptive match-making system reflecting the explicit and implicit preferences of users. *Expert Systems with Applications*, 40(4), pp.1196–1204.
- [5] Neumann, A.W. 2009. *Recommender Systems for Information Providers*. Physica-Verlag, A Springer Company.
- [6] Kelly, D. and Belkin, N. J., 2004. Display time as implicit feedback: understanding task effects. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '04*, New York, NY, USA, pp. 377–384.
- [7] Oard, D. W. and Kim, J., 2001. Modeling information content using observable behavior. *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, USA, pp.38-45.
- [8] Kelly, D. and Teevan, J., 2003. Implicit feedback for inferring user preference: a bibliography, *ACM SIGIR Forum*, 37, pp 18-28.
- [9] Guo, Q. and Agichtein, E., 2012. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. *Proceedings of the 21st international conference on World Wide Web*. pp. 569–578.
- [10] Buscher, G., White, R. W., Dumais, S. and Huang, J., 2012. Large-scale analysis of individual and task differences in search result page examination strategies. in *Proceedings of the fifth ACM international conference on web search and data mining*, pp. 373–383.
- [11] Tyler, S. K., Wang, J. and Zhang, Y., 2010. Utilizing re-finding for personalized information retrieval. *Proceedings of the 19th ACM international information and knowledge management*, pp. 1469–1472.
- [12] Precision and recall, Retrieved from [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)

- [13] Kantrowitz, M.. 2000. Stemming and its effects on TFIDF Ranking. *Proceedings of the 33rd annual international ACM SIGIR conference on Research and development in information retrieval*, 357–359.
- [14] Errami, M. 2007. ETBLAST: A web server to identify expert reviewers, appropriate journals and similar publications. *Nucleic Acids Research*, 35(S2).
- [15] Reiswig, J. 2010. Mendeley. *Journal of the Medical Library Association*. 98, (2010), 193–194.
- [16] George, T. 2005. A scalable collaborative filtering framework based on co-clustering. *Fifth IEEE International Conference on Data Mining*, pp. 625–628
- [17] Kay, J. 2006. Scrutable adaptation: Because we can and must. *Adaptive Hypermedia and Adaptive Web-Based Systems, 4th International Conference*, Dublin, Ireland, pp. 11–19.
- [18] Linden, G., Smith, B., York, J. 2003, Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing* ,7(1), 76–80.
- [19] T. Mei, B. Yang, X. Hua, L. Yang, S. Yang, and S. Li, 2007. VideoReach: an online video recommendation system. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM*. pp. 767–768.
- [20] A. Krishnakumar. 2007. Recoo: A Recommendation System for Youtube RSS Feeds. University of California, Santa Cruz, Tech. Rep.
- [21] Radev, D. R. , Qi, H., Wu, H., Fan, W. 2002. Evaluating web-based question answering systems. *Proceedings of LREC*.



## پیوست ۱- راهنمای استفاده از ژورنال یاب

### ورود عنوان مقاله

عنوان مقاله را در قسمت "عنوان مقاله" وارد می‌کنیم

#### عنوان مقاله

برجسب زنی خودکار نقش های معنایی در جملات فارسی به کمک درخت های وابستگی

### ورود چکیده مقاله

چکیده مقاله را در قسمت "چکیده مقاله" وارد می‌کنیم

#### چکیده مقاله

تشخیص خودکار واژه های دارای نقش های معنایی (همچون کنش گر، کنش پذیر، منشاء و ...) در جملات و اختصاص صحیح نقش های معنایی به آن ها توسط رایانه

## تعیین وضعیت رتبه ژورنال

با انتخاب یکی از گزینه‌ها باید وضعیت ژورنال‌های مورد نظر را انتخاب کنیم. در صورت عدم انتخاب سیستم پیام خطا می‌دهد.

### وضعیت رتبه ژورنال

فاقد درجه علمی	<input checked="" type="checkbox"/>
علمی ترویجی	<input checked="" type="checkbox"/>
علمی پژوهشی	<input checked="" type="checkbox"/>

## محل اخذ رتبه ژورنال

در این قسمت با انتخاب محل اخذ رتبه ژورنال را انتخاب می‌کنیم. محققین عزیز باید دقت داشته باشند به عنوان مثال ممکن است نشریه‌ای که دانشگاه آزاد یا حوزه علمیه رتبه علمی پژوهشی اخذ کرده باشد در وزارت علوم هیچ‌گونه رتبه‌ای نداشته باشد.

### محل اخذ رتبه ژورنال

غیره	<input checked="" type="checkbox"/>
حوزه علمیه	<input type="checkbox"/>
دانشگاه آزاد	<input type="checkbox"/>
وزارت بهداشت	<input type="checkbox"/>
وزارت علوم	<input checked="" type="checkbox"/>

## موضوع مقاله

در این قسمت کاربر می‌تواند با انتخاب موضوع محدوده نشریات مورد جستجو را محدود نماید.

### موضوع مقاله

علوم انسانی	<input checked="" type="checkbox"/>
منابع طبیعی	<input type="checkbox"/>
فنی و مهندسی	<input checked="" type="checkbox"/>
علوم کشاورزی	<input type="checkbox"/>
هنر و معماری	<input type="checkbox"/>
علوم پایه	<input checked="" type="checkbox"/>
دامپزشکی	<input type="checkbox"/>
پزشکی	<input type="checkbox"/>
غیره	<input checked="" type="checkbox"/>
عمومی	<input checked="" type="checkbox"/>

ارسال <

بعد از انجام تنظیمات مورد نظر با زدن دکمه ارسال پیشنهادات سامانه ژورنال یاب به کاربر نمایش داده می شود. در شکل زیر لیست نتایج و امکانات آن نمایش داده شده است.

جستجو در همه ...

مرتب سازی نتایج بر اساس ستون مورد نظر

نمایش ۱۰ رکورد

نام ژورنال | ضریب تاثیر ISC | دوره انتشار | وضعیت رتبه | محل اخذ رتبه | موضوع | نمایه

یژهوش و برنامه ریزی در آموزش عالی	۰.۲۵۰	فصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC, SID, magiran	بیشتر ...
پردازش علام و داده ها	۰.۱۰۰	فصلنامه	علمی پژوهشی	وزارت علوم	فنی و مهندسی	ISC, SID, Magiran, مجل اسکالر	بیشتر ...
مطالعات ترجمه	۰.۰۰۰	فصلنامه	علمی پژوهشی	وزارت علوم	علوم انسانی	ISC	بیشتر ...
شیکه	تدارد	ماهنامه	فاقد درجه علمی		فنی و مهندسی		بیشتر ...
رایانه	تدارد	ماهنامه	فاقد درجه علمی		فنی و مهندسی		بیشتر ...
علم الکترونیک و کامپیوتر	تدارد	ماهنامه	فاقد درجه علمی		فنی و مهندسی		بیشتر ...
خودرو گستر	تدارد	ماهنامه	فاقد درجه علمی		فنی و مهندسی		بیشتر ...
تدبیر	تدارد	ماهنامه	فاقد درجه علمی		علوم انسانی		بیشتر ...
بزرگراه رایانه	تدارد	ماهنامه	فاقد درجه علمی		فنی و مهندسی		بیشتر ...
مناطق ویژه اقتصادی	تدارد	ماهنامه	فاقد درجه علمی		علوم انسانی		بیشتر ...

نام ژورنال | ضریب تاثیر ISC | دوره انتشار | وضعیت رتبه | محل اخذ رتبه | موضوع | نمایه

قبلی ۱ ۲ بعدی

رفتن به صفحه بعدی نتایج

تعداد نتایج

نمایش ۱ تا ۱۰ از ۱۵ رکورد

## گزینه بیشتر

با فشردن گزینه بیشتر در لیست نتایج اطلاعات دقیق و جزئیات نشریه مورد نظر نمایش داده می‌شود.

**اطلاعات ژورنال**

**پردازش علانم و داده ها**

موضوع اصلی: فنی و مهندسی
موضوع فرعی: برقی و کامپیوتر
نمایه: ISC, SID, Magiran, گوگل اسکالر
حوزه: "زمینه‌های پردازش صوت، تصویر، متن، رمز، امنیت اطلاعات و مهندسی پزشکی"
وضعیت انتشار آخرین شماره: انتشار به موقع
تاریخ آغاز اعتبار: ۱۳۸۶/۱۲/۰۱
کد مصوب کمیسیون وزارت علوم: ۵۴۷
صاحب امتیاز: پژوهشگاه توسعه فناوری های پیشرفته خواجه نصیر الدین طوسی
سر دبیر: احمد اکبری
مدیر مسئول: جواد شیخ‌زادگان
<a href="http://jsdp.rcisp.ac.ir">http://jsdp.rcisp.ac.ir</a>
۸۸۴۴۷۱۶۳
"تهران، تقاطع استاد مطهری و سهروردی شمالی، بن بست بیسه، پلاک ۱۲"
Email : jsdp@rcdat.ir
P - ISSN: ۲۵۲۸-۴۲۰۱

**بازگشت**

ایمیل سامانه ژورنال یاب جهت پاسخ‌گویی به کاربران [journalfinder@ricest.ac.ir](mailto:journalfinder@ricest.ac.ir) می‌باشد.

## پیوست ۲- رونمایی از سامانه ژورنال یاب ایران

این سامانه در مورخ ۱۳۹۶/۹/۶ در هفته پژوهش در کارگاه آشنایی با نظام های استنادی داخلی و خارجی رونمایی

شد. در این کارگاه ضمن معرفی سامانه ژورنال یاب ایران مثال های کاربردی برای شرکت کنندگان اجرا شد و ضمن

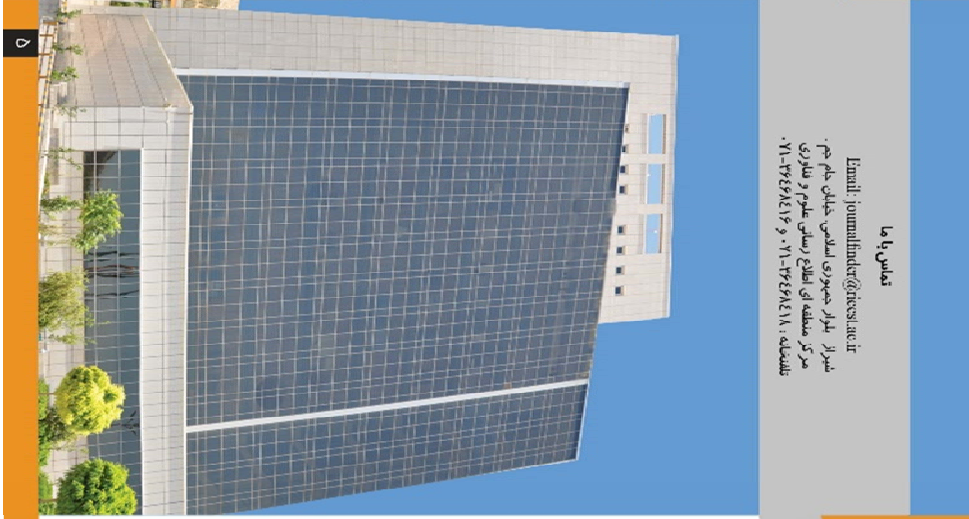
نظرسنجی به سوالات آن ها پاسخ داده شد. در ادامه بروشور سامانه ژورنال یاب آمده است.

سازمان  
RICEST  
JOURNAL ENGINEERING

جمهوری اسلامی ایران  
وزارت علوم، تحقیقات و فناوری

مرکز تحقیقاتی دانش فنی و فناوری  
وزارت علوم، تحقیقات و فناوری

http://journalfinder.ricest.ac.ir



تماس با ما  
 Email: journalfinder@ricest.ac.ir  
 شماره پست الکترونیک: journalfinder@ricest.ac.ir  
 مرکز منطقه ای اطلاع رسانی علوم و فناوری  
 تلفن: ۰۲۱-۳۲۶۹۸۱۸ و ۰۲۱-۳۲۶۹۸۱۷

این سامانه به صورت آنلاین اجرا طراحی شده و بر روی موبایل و تبلت نیز قابل استفاده است.

این سامانه به صورت آنلاین اجرا طراحی شده و بر روی موبایل و تبلت نیز قابل استفاده است.

## آشنایی با جوینگی استفاده از ژورنال های RICEST

مورفی

با کمک زدن سر روی گوگل به پستبر، اطلاعات کمپنی ژورنال قابل مشاهده خواهد بود. مواردی مانند حوزه موضوعی ژورنال و وضعیت انتشار آخرین شماره در تصمیم گیری برای انتخاب ژورنال هدف جهت چاپ و انتشار مقاله کمک نماید.

با کمک زدن سر روی گوگل به پستبر، اطلاعات کمپنی ژورنال قابل مشاهده خواهد بود. مواردی مانند حوزه موضوعی ژورنال و وضعیت انتشار آخرین شماره در تصمیم گیری برای انتخاب ژورنال هدف جهت چاپ و انتشار مقاله کمک نماید.

برای استفاده از این سامانه در ابتدا عنوان مقاله و یکجمله مقاله را در جعبه مربوطه وارد نمایید سپس نتایجش را نامعلوم داده بر روی دکمه ارسال کلیک نمایید.

پس از چاره‌اندیشی در خواست شما فهرست بهترین پیشنهادها سامانه به شما نمایش داده می شود. شما می توانید این پیشنهادها را بر اساس صریح تاخیر، دوره انتشار، وضعیت رتبه، محل اکتساب رتبه موضوعی و پایگاه های نمایه کننده مرتب نمایید.

ردیف	موضوعی	سال	نوع مقاله	مجله	سال انتشار
1	موضوعی	سال	نوع مقاله	مجله	سال انتشار
2	موضوعی	سال	نوع مقاله	مجله	سال انتشار
3	موضوعی	سال	نوع مقاله	مجله	سال انتشار
4	موضوعی	سال	نوع مقاله	مجله	سال انتشار
5	موضوعی	سال	نوع مقاله	مجله	سال انتشار
6	موضوعی	سال	نوع مقاله	مجله	سال انتشار
7	موضوعی	سال	نوع مقاله	مجله	سال انتشار
8	موضوعی	سال	نوع مقاله	مجله	سال انتشار
9	موضوعی	سال	نوع مقاله	مجله	سال انتشار
10	موضوعی	سال	نوع مقاله	مجله	سال انتشار

ایجاد سامانه ای برای کمک به پژوهشگران در یافتن ژورنال مناسب جهت انتشار مقاله خود را اهداف این سامانه می باشد. در سال های اخیر ناشران معروف دنیا مانند Springer و Elsevier با ارائه چنین سامانه هایی کمک فراوانی به توسعه کار مقالات کرده اند. تاکنون چنین سامانه ای برای ژورنال های علمی فارسی نمونه ای نداشته است و پژوهشگران فارسی زبان از این نظر محرومیت داشته اند.

با توجه به عنای بالای پایگاه داده RICEST در حالت کار ۳۳۰۰ عنوان ژورنال فارسی رتبه کار سامانه ژورنال باب بعد از تحقیقات علمی توسط اعضای هیئت علمی و کارکنان پژوهشی گروه علمی و عملیات سیستم ها RICEST پیغام سازی شد. الگوریتم استفاده شده در این سامانه هیچ نمونه ای در دنیا نداشته و اساساً ناشران معروف دنیا نیز الگوریتم های مورد استفاده خود را عمومی کرده اند.

برای ارزیابی این سامانه، تویم توسعه دهنده مجموعه ای منتخب از ۴۰۰ مقاله از پایگاه داده مقالات فارسی RICEST را مورد تست قرار داده اند و دقت دست آمده از آن را حدود ۸۰٪ اعلام کرده اند که می تواند باعث اطمینان بعضی نتایج دست آمده از این سامانه شود.

### اهداف

- سامانه ای برای کمک به پژوهشگران جهت یافتن بهترین ژورنال برای انتشار تحقیقات خود
- پیش بینی میزان شباهت موضوع مقاله پژوهشگران با ژورنال های کشور
- ارائه مشخصات ژورنال ها و ضریب تاثیر آن ها