



## محتوای الکترونیکی زبان فارسی و ارتباط آن با جویشگر بومی

رضا داوطلب<sup>۱</sup>، محمد آزاد نیا<sup>۲</sup>

<sup>۱</sup> دانشجوی دکتری، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، تبریز،  
r.davtaleb@tabrizu.ac.ir

<sup>۲</sup> مربی، پژوهشکده فناوری اطلاعات، مرکز تحقیقات مخابرات، تهران  
azadnia@itrc.ac.ir

### چکیده

محتوای الکترونیکی یکی از مؤلفه‌های زیرساختی حوزه فناوری اطلاعات بوده و با توجه به اهمیتی که در افزایش کیفیت سرویس‌های نرم‌افزاری دارد، همواره مورد توجه بوده است. با گسترش اینترنت و سرویس‌های مبتنی بر آن، محتوای الکترونیکی نقش پررنگ‌تری در این حوزه به خود گرفته است. در این نوشتار قصد داریم تا با بررسی ابعاد مختلف محتوای الکترونیکی، جایگاه آن را در ارتباط با طرح جویشگر تبیین نموده و وضعیت محتوای الکترونیکی زبان فارسی را مورد بررسی قرار دهیم. برای اینکار ابتدا به مرور مفاهیم اولیه محتوای الکترونیکی پرداخته و پس از آن ارتباط جویشگر و محتوا را بررسی خواهیم نمود. در نهایت نیز مهمترین مشکلات موجود در حوزه محتوای الکترونیکی را بیان خواهیم نمود.

### کلمات کلیدی

محتوای الکترونیکی، جویشگر بومی، مدل‌های نشر داده، تولید محتوا، چالش‌های تولید محتوا، شبکه‌های اجتماعی.

### ۱- مقدمه

در اصطلاح فناوری اطلاعات، به اطلاعات و دانشی که در بسترهای الکترونیکی قابل نشر هستند را محتوای الکترونیکی گویند. با توجه به حجم بسیار بالای اطلاعاتی که امروز تولید و عرضه می‌شود، ارائه و نشر این اطلاعات بدون در نظر گرفتن اینترنت و شبکه‌ی جهانی وب، غیرقابل تصور است. بنابراین، منظور از محتوای الکترونیک در این نوشتار، در حقیقت محتوای الکترونیکی قرار گرفته در بستر وب است [۲].

اطلاعاتی که توسط مردم سرزمینی در بستر اینترنت جاری می‌شود، برخاسته از فرهنگ و بوم آن سرزمین است. هرچه این اطلاعات و محتوای ارائه شده غنی‌تر باشد، ملت آن سرزمین به فرهنگ و بوم خود وابستگی بیشتری داشته و مخاطراتی مانند تهاجم فرهنگی در آن جامعه کمتر احساس خواهد شد. همچنین، ملتی که محتوای بیشتر و با کیفیتی در بستر اینترنت منتشر می‌کند، کاربران بیشتری به سمت خود و فرهنگ و زبان سرزمین خود جذب می‌نماید.

توسعه جامعه اطلاعاتی، اقتصاد دانایی محور\* را شکل می‌دهد. اقتصاد دانایی محور، اقتصادی است که به طور مستقیم بر مبنای تولید، توزیع و استفاده از دانش و اطلاعات است. واژه اقتصاد دانایی محور نتیجه شناخت بهتر از نقش دانش و فناوری در رشد اقتصادی است. دانش از مهمترین عوامل رشد اقتصادی محسوب می‌شود، از اینرو اقتصاددانان به دنبال کشف راه‌هایی برای وارد کردن عامل دانش و فناوری به مدل‌ها و تئوری‌های رشد اقتصادی می‌باشند. از مهمترین اهداف و نتایج اقتصاد دانایی محور افزایش رشد اقتصادی، ایجاد محیط مناسب برای سرمایه‌گذاری، ایجاد اشتغال، ارتقای بهره‌وری و توسعه تجارت خارجی می‌باشد [۳]. بنابراین بیشتر کشورهای پیشرو (توسعه یافته و در حال توسعه) توسعه فناوری اطلاعات را عامل و فرصتی مهم برای رشد و توسعه اقتصادی می‌انگارند [۴].

### ۲- تولید و نشر محتوای الکترونیکی

انتشار داده‌ها در بستر وب را می‌توان بر اساس نوع تولید و نوع انتشار، در چند دسته مختلف قرار داد. این مدل‌ها در دوره‌های زمانی مختلف حیات شبکه جهانی وب و با توجه به مقتضیات زمانی هر دوره، به وجود آمده‌اند که در ادامه برخی از این مدل‌ها معرفی می‌گردند.

#### ۲-۱- مدل اول: انتشار یکطرفه

در دوره نخست پیدایش اینترنت که از ابتدا تا حدود سال ۲۰۰۰ واقع شده، عصر دسترسی می‌گوئیم. موضوع کانونی این دوره ضریب نفوذ اینترنت، سرویس‌های مختلف اینترنتی و بعضی زیر ساخت‌های اینترنتی و وضعیت کانال‌های دسترسی به اینترنت و روش‌ها و ابزارهایی بوده است که امکان اتصال کاربر را به این فضا فراهم می‌آورد. الگوی ارتباطی این دوره، مدل کلاسیک ارتباط رسانه‌ای به معنای فرستنده‌های اندک و گیرندگان انبوه بود که در تمام رسانه‌های دیگر مانند روزنامه، رادیو، تلویزیون، صنعت سینما، کتاب و رسانه‌های مشابه وجود داشت [۴]. این مدل انتشار برای تولید کنندگان محتوای هزینه‌بر بوده و تولید کننده محتوای بایستی کسب و کار مناسب با خود را در این زمینه داشته باشد. بنابراین کاربر ممکن است در قبال دریافت محتوای مورد نیاز خود مجبور به پرداخت هزینه باشد و یا اینکه تولید کننده با انگیزه‌ای غیر از کسب در آمد، اقدام به نشر اطلاعات نماید.

این مدل امروزه بیشتر توسط سایتهای خبری و سایت‌هایی که در قبال محتوای خود کسب درآمد می‌کنند، مورد استفاده قرار می‌گیرد. در شکل ۱ شمای کلی این مدل نشان داده شده است. همانطور که ملاحظه می‌گردد، سرویس ارائه شده نیز توسط سازمان یا شرکت تولید کننده محتوا توسعه داده می‌شود.

البته مدل دیگری هم در این ارتباط وجود دارد که از ترکیبی از دو مدل معرفی شده قبلی استفاده می‌نماید. به عنوان مثال گوگل در سرویس نقشه خود هم از اطلاعات وارد شده توسط کاربران استفاده می‌کند و هم اینکه اینکه رفتار آنان را مورد تحلیل قرار داده و سرویس کاملتری ارائه می‌دهد. مزیت عمده این روش نسبت به مدل‌های قبلی این است که در این مدل کاربر خود را مجبور به مشارکت در امر تهیه داده نمی‌بیند و احساس دریافت رایگان سرویس و محتوای ارائه شده را دارد. از معایب این روش نیز می‌توان به استفاده از روش‌ها و الگوریتم‌های پیچیده برای تحلیل رفتار کاربر و استخراج اطلاعات از آن اشاره نمود که باعث پیچیدگی زیاد پیاده‌سازی چنین سرویس‌هایی می‌شود. البته باید گفت که متأسفانه، چنین شرکت‌هایی علاوه بر استفاده از اطلاعات کاربران خود، این اطلاعات را در ازای دریافت مبالغ هنگفتی در اختیار شرکت‌ها و سازمان‌های دیگر قرار می‌دهند که این امر می‌تواند مشکلاتی را برای کاربران این شرکت‌ها به وجود آورد.

امروزه بیشتر شرکت‌های بزرگ و شبکه‌های اجتماعی مطرح جهانی به سمت استفاده از اطلاعات کاربران حرکت نموده‌اند. اگر وضعیت کنونی این شبکه‌ها را که روزانه میلیاردها نفر کاربر دارد در نظر بگیریم، هیچ کدام، اطلاعاتی به عنوان مدیران پلتفرم‌ها تولید نمی‌کنند، بلکه کاربران این داده‌ها را روی آن بارگذاری می‌کنند. براساس آخرین آمار، داده‌های موجود در فضای مجازی به طور متوسط هر سه سال دو برابر می‌شود. این بدان معناست که بیش از ۵۰ درصد اطلاعاتی که هم‌اکنون در فضای مجازی در اختیارمان قرار دارد مربوط به ۳-۲ سال اخیر است [۵] که سرعت رشد تولید محتوا را در این فضا بیان می‌کند. با توجه به این موضوع، شاهد شکل‌گیری بانک‌ها یا پایگاه‌های اطلاعاتی هستتیم که در بحث‌های فنی به عنوان کوهی از اطلاعات BIG-DATA در نظر گرفته می‌شود. بنابراین کسانی بردگان این بازی و این دوران هستند که توانایی کاوش در داده‌های عظیم و تولید دانش از آن را داشته باشند [۶].

### ۳- حوزه‌های کاربردی محتوای الکترونیک

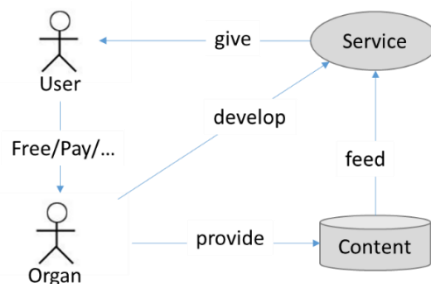
آنچه که از کاربرد محتوا انتظار می‌رود می‌تواند جنبه‌های متعددی مانند: تحقیق، توسعه/بهبود بخشی/بهبودسازی، ترویج / آموزش/تبلیغ، توجیه، ارائه طریق (راهبرد و راهکار)، ارائه ایده، برنامه‌ریزی، تحلیل/ارزیابی/تضمین را پوشش‌اند. عبارتی دیگر، یک محتوا می‌تواند برای تحقیق، توسعه و غیره بکار گرفته شود. برای هر یک از این موارد می‌توان ذکر کرد که چه مفاهیمی لازم است تا در محتوا لحاظ گردند. بعنوان مثال، محتوایی که جهت تحقیق استفاده می‌شود، باید پاسخگوی موضوع تحقیق، پدیده‌ها و رویدادهای تحقیق، کیفیت و کمیت تحقیق باشد. از طرف دیگر، جنبه‌هایی که در توسعه لازم است، مدنظر قرار گیرند عبارتند از: فرآیند، برنامه، ضروریات، ساختار/تشکیلات، اعتبار، هدف، منبع، کیفیت و کارآمدی/کارایی. بدین ترتیب برای تمام کارکردهای محتوا می‌توان جنبه‌های ضروری را متصور شد [۱].

### ۴- نقش جویشگرها در محتوا

در تمامی مدل‌های معرفی شده جویشگرها نقش انکار ناپذیری در نشر محتوا دارد، زیرا نزدیک ۸۰ درصد کاربران فضای مجازی از طریق یکی از جویشگرها جستجو وارد محیط مجازی می‌شوند [۷]. در مدل انتشار یک‌طرفه، نقش موتور جستجو بیشتر معرفی کننده محتوای تولید شده توسط شرکت‌های تولید کننده محتوا است که در این حالت، معرفی محتوای با کیفیت تولید شده توسط این شرکت‌ها می‌تواند باعث رونق کسب و کار در این زمینه شده و محرکی برای صنعت تولید محتوای کشور تلقی گردد.

در حالت انتشار تعاملی، با توجه به اینکه همه کاربران به نوعی با موتورهای جستجو در ارتباط هستند، می‌توان از این ارتباط برای تحلیل رفتار کاربران استفاده نمود و اطلاعات ارزشمندی در این خصوص استخراج نمود. در این حالت جویشگر خود در جایگاه یک تولید کننده محتوای ظاهر شده و علاوه بر استخراج اطلاعات مورد نیاز خود از این داده‌ها، می‌تواند اطلاعات مورد نیاز سایر شرکت‌ها و نهادهای مرتبط با شبکه‌ی وب را استخراج نماید. این اهمیت تا حدی است که ارائه محتوای با کیفیت بدون وجود جویشگر ارزشی نخواهد داشت [۵].

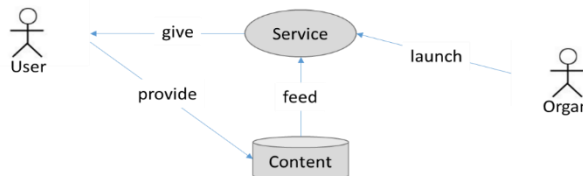
از طرف دیگر، بخش زیادی از محتوای با ارزش داخل کشور که بین کاربران طرفداران زیادی هم دارند، معمولاً در اختیار نهادهای حکومتی و دولتی هستند. سهم بسیار زیادی از این محتوا، بنا به محدودیت‌های خاص متولیان خود، قابل ارائه به سرویس‌های جویشگر خارجی مانند گوگل و سایر جویشگرها نیستند. بنابراین امکان ارائه این محتوا به کاربران وجود نخواهد داشت. از مهمترین این منابع می‌توان به مجموعه‌های محتوایی تحت نظر شورای عالی اطلاع رسانی و سازمان تبلیغات اسلامی، سازمان اسناد و کتابخانه ملی، مجموعه پایان نامه‌های کشور که در اختیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) و محتوای تولید شده در دانشگاه‌ها و مراکز وزارت علوم و وزارت بهداشت و درمان اشاره کرد. توجه داشته باشید که با توجه به اینکه ارزش بسیاری از اطلاعات وابسته به زمان است، این محتوا مانند منبع غذایی با ارزشی است که



شکل ۱: مدل انتشار یک طرفه محتوا در بستر وب

### ۲-۲- مدل دوم: انتشار تعاملی

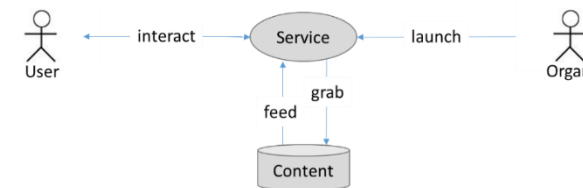
از سال ۲۰۰۰ به بعد که به تعبیری آغاز دوره وب ۲ است، با ظهور بلاگ‌ها شروع می‌شود. این دوره را می‌توانیم عصر شبکه‌های اجتماعی و انتشار تعاملی نامگذاری کنیم. در این دوره با ظهور فناوری وب ۲ بویژه پدیده یا فناوری شبکه‌های اجتماعی مدل ارتباطی قبل تا حدی تغییر کرده و کاربران تبدیل به تولیدکننده و مصرف‌کنندگان اطلاعات شدند. بنابراین مدل ارتباط رسانه‌ای کلاسیک به مدل "فرستنده و گیرنده تعاملی" تبدیل شد. بنابراین نقش گیرنده از بازخورد دهنده فراتر رفته و تبدیل به تولیدکننده محتوا شده است. علاوه بر آن، از سال ۲۰۱۰ به بعد شاهد پدیده دیگری هستتیم که می‌توانیم از آن به عنوان عصر محتوا یاد کنیم و پیش‌بینی می‌شود حداقل تا پنج سال آینده نیز ادامه داشته باشد. در این دوره مفاهیم جدیدی مانند سکوی انتشار یا پلتفرم به وجود آمد [۵]. در این مفهوم، برخلاف دوره‌های پیشین، کاربران، مهمترین نقش را در تولید محتوا دارند. در این مدل پلتفرم‌های قدرتمندی به عنوان سکوهایی بارگذاری اطلاعات ایجاد شده و با هجوم شبکه‌های کاربران به سمت این پلتفرم‌ها، داده‌ها، اطلاعات و محتواها توسط کاربران روی آنها بارگذاری می‌شود. هر چند چگونگی گرایش کاربران به این پلتفرم‌ها نیز پیچیدگی خاصی دارد که لزوماً به معنای برتری فناوری نیست [۴]. در این مدل از ارتباط، محتوای مربوط به کاربر از دو کانال متفاوت جمع آوری می‌گردد. در حالت اول، کاربران از سرویس‌های ارائه شده استفاده کرده و در صورت تمایل اقدام به غنی‌تر کردن محتوای موجود در سایت می‌پردازند که از جمله این سرویس‌های می‌توان به ویکی‌پدیا و یا سیستم‌های پرسش و پاسخ اشاره نمود. در شکل ۲ یک شمای کلی این مدل نشان داده شده است.



شکل ۲: مدل انتشار تعاملی - جمع آوری اطلاعات توسط کاربران

همانطور که ملاحظه می‌شود. در این مدل کاربر از سرویس ارائه شده استفاده کرده و در مواقعی، به تکمیل نمودن محتوای موجود می‌پردازد. البته در این مدل، خود شرکت متصدی نیز در تهیه محتوای مورد نیاز اولیه برای اجرای سرویس مشارکت دارد که این محتوای اولیه لازمه به اجرا در آمدن چنین سرویس‌هایی خواهد بود. در حالت دوم کاربر تنها از سرویس‌های ارائه شده استفاده کرده و شرکت متصدی، با تحلیل رفتار کاربر و اطلاعاتی که از این طریق بدست می‌آورد، اطلاعات خود را به روز رسانی می‌کند و یا از این اطلاعات در سرویس‌های دیگر خود استفاده می‌کند. به عنوان مثال سرویس ترافیک گوگل از جمله این سرویس‌هاست که با تحلیل رفتار و موقعیت کاربران به جمع آوری داده می‌پردازد.

در شکل ۳ شمای کلی مدل ذکر شده ملاحظه می‌گردد. همانطور که در این شکل نشان داده شده است، کاربر، تنها با سرویس در تعامل بوده و کمک مستقیمی در غنی‌سازی محتوای مربوط ندارد.

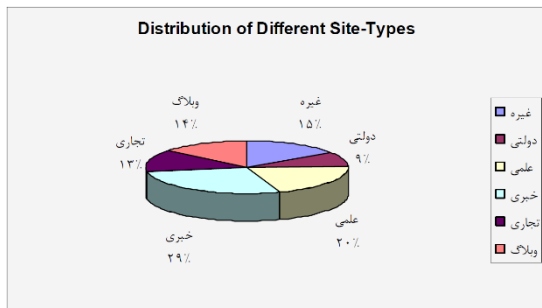


شکل ۳: مدل انتشار تعاملی - جمع آوری اطلاعات با استفاده از تحلیل رفتار کاربر

## ۶- بررسی وضعیت موجود در کشور

در بررسی وضعیت محتوای فارسی و وب کشور، ابعاد مختلفی بایستی مورد توجه قرار گیرد که از مهمترین آن می‌توان به کشور میزبان سایت‌های ارائه دهنده محتوا و دسته‌های محتوایی مورد توجه کاربران اشاره نمود. کشور میزبان محتوا از آن جهت مورد توجه قرار می‌گیرد که هم از نظر میزان ترافیک داخل به خارج کشور و هم از نظر امنیتی از اهمیت بالایی برخوردار است. طبق گزارشی که توسط معاونت توسعه و مدیریت شبکه‌های ملی اطلاعات ارائه شده است، خوشبختانه با تمهیدات پیش بینی شده در وزارت ارتباطات و فناوری اطلاعات، تعداد زیادی از سایت‌های پربازدید کشور در حال انتقال به سرورهای داخل کشور بوده و نسبت میزبانی وبسایت‌های پربازدید داخل به خارج، از حدود ۵۰ در سال ۹۳ به ۶۳ درصد در آبان ماه سال جاری رسیده است [۱۰].

از منظر دسته‌های اطلاعاتی پرکاربرد کاربران ایرانی؛ بر اساس گزارشی که در سال ۸۵ ارائه شده است [۱۱]، سهم زیادی از سایت‌های ایرانی را وبسایت‌های خبری تشکیل می‌دهند. داده‌های مورد استفاده در این پژوهش، اطلاعات مربوط به ۲ میلیون صفحه خزش شده فارسی است. طبقه‌بندی موضوعی وبسایت‌های خزش شده، در شکل ۴ قابل مشاهده است.



شکل ۴: توزیع سایت‌های فارسی خزش شده بر اساس طبقه‌بندی موضوعی [۱۱]

در گزارش دیگری که توسط اثنی‌عشری و همکاران [۹] ارائه شده است، برای پایش وضعیت موجود کشور در زمینه محتوا، با توجه به بررسی وبسایت‌های پرمخاطب و بر اساس دسته‌بندی انجام شده؛ داده‌های پرکاربرد کاربران ایرانی در سه کلاس مختلف دسته‌بندی شده‌اند. به این منظور داده‌هایی مانند داده‌های مکانی، آگاهی عمومی و خبر پرکاربردتر از سایر موارد هستند. در سطح بعدی اهمیت، داده‌هایی نظیر محتوای آموزشی، محتوای پژوهشی، اطلاعات اماکن پر مراجعه، آرشیو فیلم، کارتون، سریال، محتوای مذهبی و آرشیو آگهی‌های استخدام قرار گرفته و در دسته کم اهمیت‌ترین نیز داده‌هایی مانند اطلاعات املاک، اطلاعات حمل و نقل و اطلاعات کالاهای فروشی مستعمل قرار دارند.

## ۷- مشکلات موجود در حوزه محتوای الکترونیک زبان فارسی

علی‌رغم پیشینه تاریخی بسیار غنی در زمینه تولید محتوای زبان فارسی، امروزه از جایگاه محکمی در این حوزه برخوردار نیستیم. می‌توان گفت در این عرصه میدان را به محتوای بیرونی واگذار کردیم. در صنعت بازی، وب گاه‌ها، نشر و بسیاری از حوزه‌های دیگر بیشتر مصرف‌کننده هستیم و یا اینکه از ترجمه آثار خارجی استفاده می‌کنیم. البته در سال‌های اخیر قوانین بسیاری در رابطه با تولید و نشر محتوای داخلی وضع شده ولی واقعیت این است که این قوانین روی کاغذ می‌ماند و گرچه مقررات تدوین و ابلاغ می‌شود اما دستگاه‌های کشور به جای اجرای آموزه دینی «زکات العلم نشره» مایل نیستند یا توانایی و اجازه آن را ندارند یا تکنیک مربوطه را نمی‌دانند تا محتواهایی که تولید می‌کنند را در وب بارگذاری و منتشر کنند. در راستای این موضوع، حدود هفت سال پیش بود که کتابخانه ملی ایران، کتابخانه آستان قدس رضوی، شورای عالی اطلاع‌رسانی، کتابخانه مرکزی دانشگاه تهران و چند نهاد دست اندرکار دیگر، یک کنسرسیوم تولید محتوا در فضای مجازی تشکیل داده‌اند که یکی از مهمترین خروجی‌های آن، توسعه یک وبسایت متمرکز برای دسترسی مناسب کاربران به محتوای ارائه شده توسط اعضای کنسرسیوم است. با توجه به جمع‌آوری مجموعه غنی از محتوای الکترونیکی فارسی در

در زمان قحطی، افراد جامعه امکان بهرمندی از آن را نداشته و پس از طی مدت کوتاهی فاسد خواهد شد. بنابراین نیاز سرویسی داریم تا ارتباط میان کاربران و محتوای ارائه شده را به نحو مطلوبی برقرار نماید و بهترین راه ارائه چنین سرویسی، جویشگر خواهد بود. حال اگر این جویشگر داخلی باشد، افراد و نهادهای بیشتری به آن اعتماد کرده و محتوای گرانمای خود را در اختیار آن خواهند گذاشت.

از طرف دیگر باید توجه داشت که با توجه به رونق کسب و کارهای مبتنی بر فناوری اطلاعات در کشورهای توسعه یافته، افراد و شرکتهای زیادی راغب به فعالیت در این حوزه هستند. به عنوان مثال شرکت‌هایی با هدف جمع‌آوری داده و فروش آن به متقاضیان آن به وجود آمده‌اند. این شرکت‌ها با نام تجاری داده شناخته شده و جویشگرها، اصلی‌ترین مشتری این شرکت‌ها هستند. تجزیه‌کنندگان داده، داده‌های اطلاعاتی مناسب و پرمخاطب را جمع‌آوری نموده و سعی می‌کنند آنها را بروز نگاه دارند. در حقیقت، کسب و کار این شرکت‌ها بر پایه ارائه اطلاعات بروز و مناسب به مخاطبین است و اگر نتوانند اطلاعات خود را بروز کنند، کسب و کار خود را از دست خواهند داد.

## ۵- تعامل جویشگران با تجزیه‌کنندگان داده

همکاری بین جویشگران و تجزیه‌کنندگان از آنجا شکل می‌گیرد که اتکالی صرف به داده‌های خزش شده، می‌تواند دقت اطلاعات ارائه شده را در جویشگر کاهش دهد. برای حل این مشکل و افزایش دقت داده‌هایی از این دست، جویشگرهای مطرح نظیر گوگل و بینگ سعی نموده‌اند قراردادهایی را با تجزیه‌کنندگان داده محلی امضاء نمایند تا بدین ترتیب بتوانند به اطلاعات بروز، دقیق و مناسب دست یابند. این جویشگران داده‌های خزش شده خود را با داده‌های موجود در پایگاه داده این شرکت‌ها مقایسه و اعتبارسنجی می‌نمایند. بدین ترتیب، این جویشگرها می‌توانند مطمئن باشند که اطلاعات بروز و مناسبی را در اختیار مخاطبین خود قرار می‌دهند [۸]. این مسأله از یک طرف باعث افزایش کاربران جویشگر شده و از طرف دیگر، با توجه به قراردادی که مابین جویشگر و تجزیه‌کننده داده وجود دارد، موجب رونق کسب و کار تجزیه‌کننده داده خواهد شد [۹].

با توجه به آنچه که بیان شد، وجود تجزیه‌کنندگان داده محلی از یک سو و ارتباط اطلاعاتی آنها با جویشگرها از سوی دیگر، موجب پیدایش چرخه‌ای از اطلاعات در هر کشور می‌شود. در مرجع [۹]، Error! Reference source not found. نحوه تعامل تجزیه‌کنندگان داده با جویشگران و ارائه‌کنندگان سرویسهای مبتنی بر محتوا، در آمریکا، آلمان و چند کشور دیگر به تفصیل بررسی داده شده است که این مطالعات بر اساس تحقیقات آقای دیوید میهم انجام شده است.

اما در داخل کشور برای شکل‌گیری اکوسامانه‌ی اطلاعاتی جویشگر راه زیادی در پیش داریم. در این راه چالش‌های بزرگی در مقابل ما قرار دارد که بایستی راهکارهای مناسبی برای مقابله با آنها پیش‌بینی شوند که از آن جمله می‌توان به موارد زیر اشاره نمود [۹].

**عدم تمرکز جویشگرهای ایرانی بر دریافت اطلاعات از تجزیه‌کنندگان داده در سطح کشور:** جویشگرهای ایرانی، نظیر یوز و پارسی‌جو، هنوز در ابتدای راه هستند و خدمات ارائه شده توسط آنها به بلوغ کافی نرسیده است. به همین دلیل، تمرکز فعلی این جویشگرها بیشتر بر افزایش پوشش دهی صفحات فارسی و نیز افزایش دقت جستجو در میان صفحات و اسناد نمایه شده است. این مسأله باعث شده است که این جویشگرها کمتر به دنبال برقراری ارتباط با تجزیه‌کنندگان داده و دریافت داده‌های مناسب از آنها باشند.

**عدم وجود تجزیه‌کنندگان قوی در سطح کشور:** متأسفانه با توجه به نوب بودن صنعت IT در ایران، هنوز تجزیه‌کنندگان داده‌ای که بتوانند از لحاظ حجم داده، کیفیت داده، و مهم‌تر از همه بروز بودن داده‌های تجزیه‌کننده با نمونه‌های مشابه خارجی برابری نمایند در سطح ایران دیده نمی‌شود. بدیهی است که جویشگرها نمی‌توانند به داده‌هایی که بروز بودن آنها تضمین نشده است اطمینان کنند و از آنها برای ارائه خدمت به کاربران خود استفاده نمایند.

**عدم همکاری تجزیه‌کنندگان داده با جویشگرها:** در معدود مواردی که تجزیه‌کنندگان داده نسبتاً مناسبی در سطح کشور وجود دارد، که به عنوان نمونه می‌توان به اطلاعات تماس ۱۱۸، مقالات موجود در سیویلیکا و نقشه‌های سازمان نقشه‌برداری ایران و شهرداری‌ها اشاره نمود، این تجزیه‌کنندگان داده تمایلی به ارائه اطلاعات خود به جویشگرهای ایرانی نشان نداده‌اند و تمامی تلاش‌ها در این حوزه تا کنون با شکست مواجه شده است.

<sup>9</sup> David Mihm

<sup>9</sup> Data Aggregators

### ۷-۳- مشکلاتی که بر عدم استفاده از محتوای الکترونیکی موجود دلالت دارند

- بازاریابی و تبلیغات نامناسب
- سرویس‌های نامناسب

با توجه به مشکلات ذکر شده، می‌توان راهکارهای متفاوتی برای برطرف نمودن این مشکلات ارائه داد که البته همگی این راهکارها در ارتباط با بحث جویش و به طور خاص طرح جویشگر بومی نبوده و متولیان مختلفی برای آن وجود دارد. بخشی از این مشکلات و چالش‌ها که با علامت **♦** مشخص شده‌اند را می‌توان در قالب طرح جویشگر مطرح نموده و راهکارهای مربوط به آن را ارائه نمود.

### ۸- نتیجه گیری

همانطور که گفته شد، وجود محتوای غنی از مهمترین موارد حفظ فرهنگ و بوم هر کشوری در مقابل بیگانگان بوده و توسعه محتوا، ضمن گسترش فرهنگ و زبان یک سرزمین، تاثیر بسزایی در خودباوری و توسعه پایدار آن سرزمین دارد. از طرف دیگر، دسترسی به محتوای الکترونیکی غنی، از الزامات اولیه هر جویشگری بوده و بسیاری از سرویس‌های پایه طرح جویشگر، وابستگی مستقیم به محتوای الکترونیکی دارند. از این رو، توجه، توسعه و حمایت از محتوای الکترونیکی فارسی، اهمیت بالایی دارد. علاوه بر آن پیشنهاد می‌گردد، پروژه‌هایی برای تبدیل محتوای غیر الکترونیکی به محتوای الکترونیکی تعریف شده و دسترسی لازم جویشگران داخلی به محتوای موجود فراهم گردد تا همانطور که در اسناد بالا دستی کشور مانند سند چشم انداز ۲۰ ساله و برنامه ششم توسعه کشور آمده است، زمینه دسترسی عموم مردم به محتوای پاک فراهم گردد.

### ۹- مراجع

- [1] م. خراط، "ارائه چارچوب هستان‌نگاری برای بازنمایی جنبه‌های مختلف پردازش و حوزه‌های کاربری محتوای الکترونیکی"، گزارش فنی پژوهشکده فناوری اطلاعات مرکز مخابرات ایران، ۱۳۸۷.
- [2] P. Arzberger, P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, et al., "An international framework to promote access to data," *SCIENCE-NEW YORK THEN WASHINGTON*, pp. 1777-1778, 2004.
- [3] ک. بدیع، م. خراط، م. ط. محمودی، م. ا. میریان، ط. میرسعیدقاسمی، س. بابازاده و همکاران، "طرح راهبردی توسعه محتوای دیجیتال در کشور: مطالعه و بررسی فعالیت‌های تحقیقاتی سازمان توسعه و همکاری‌های اقتصادی"، (OECD)، گزارش فنی پژوهشکده فناوری اطلاعات مرکز مخابرات ایران، vol. ۷ شماره گزارش ۱۵-۱-۲۰۱-۱۳۸۵.
- [4] م. فاطمی، "ضرورتی به نام محتوا"، ماهنامه تحلیلگران عصر اطلاعات، ۱۳۸۹.
- [5] A. Halavais, *Search engine society*: John Wiley & Sons, 2013.
- [6] T. Heath and C. Bizer, "Linked data: Evolving the web into a global data space," *Synthesis lectures on the semantic web: theory and technology*, vol. 1, pp. 1-136, 2011.
- [7] Y. Chen, G. Y. Jeon, and Y.-M. Kim, "A day without a search engine: an experimental study of online and offline searches," *Experimental Economics*, vol. 17, pp. 512-536, 2014.
- [8] A. Brueckman, "How to Improve Your Local SEO with Data Aggregators," 2014.
- [9] م. اثی‌عشری، م. عظیم‌زاده و م. محمودی، "اکوسمانه فضای اطلاعاتی جویشگر بومی"، گزارش فنی طرح جویشگر، ۱۳۹۴.
- [10] معاونت توسعه و مدیریت شبکه‌ی ملی اطلاعات، "گزارش تفصیلی: پایش وضعیت میزبانی ۵۰۰ وب‌گاه پربازدید کاربران داخلی"، سازمان ارتباطات و فناوری اطلاعات کشور، ۱۳۹۴.
- [11] ا. کیهانی پور، ع. محمدزاد بیدی، م. محمودی و م. آزادانیا، "ارزیابی محتوای وب ایران از منظر دولت الکترونیک"، ارائه شده در دوازدهمین کنفرانس سالانه انجمن کامپیوتر ایران، ۱۳۸۵.
- [12] "درگاه پایش جامعه اطلاعاتی جمهوری اسلامی ایران *The Official Portal of Measuring Information Society of Iran*, vol. <http://mis.iran.ir/>.

این کنسرسیوم، در صورت برقراری ارتباط مناسب، میان جویشگرهای بومی و این مجموعه، دسترسی کاربران به محتوای مذکور تسهیل گردیده و استفاده از این محتوای ارزشمند به نحوه چشم‌گیری افزایش می‌یابد.

البته با توجه به گستردگی بیش از حد مساله محتوا، به نظر می‌رسد در این حوزه شاخص‌های دقیقی برای بررسی و پایش وضعیت محتوای الکترونیکی کشور مورد نیاز است. در آمار ارایه شده در "درگاه پایش جامعه اطلاعاتی جمهوری اسلامی ایران" [۱۲] در بخش "شاخص‌های فاوای کشور- محتوای الکترونیکی" <sup>۱۲</sup>، ۱۲ شاخص زیر مورد پایش قرار گرفته که تنها برای ۶ مورد از این شاخص‌ها وضعیت عملکردی بیان شده است:

- نسبت کتابخانه‌های عمومی با دسترسی به اینترنت پهن‌بند (۹۶،۲۸ درصد)
- نسبت کتابخانه‌های عمومی که دارای درگاه (پرتال) امن و استاندارد در ارائه محتوا هستند (۹۵،۹۷ درصد)
- نسبت آثار الکترونیکی رسمی چاپ و نشرشده به تفکیک نوع اثر (نامعلوم)
- نسبت موزه‌ها با دسترسی به اینترنت پهن‌بند (۶۳ درصد)
- نسبت موزه‌هایی که دارای درگاه (پرتال) امن و استاندارد در ارائه محتوا هستند (نامعلوم)
- نسبت سازمان‌های آرشیو ملی با دسترسی به اینترنت پهن‌بند (نامعلوم)
- نسبت سازمان‌های آرشیو ملی که دارای درگاه (پرتال) امن و استاندارد در ارائه محتوا هستند (نامعلوم)
- نسبت اسنادی در آرشیوهای ملی که دیجیتال شده‌اند (۲۰ درصد)
- نسبت اسناد دیجیتال‌شده‌ای که به شکل عمومی در دسترس آنلاین هستند (نامعلوم)
- نسبت صفحات وب به زبان فارسی (۰،۹ درصد)
- سهم مقالات فارسی در ویکیپدیا به کل مقالات (۳۰۰ هزار)
- تعداد ثبت نام دامنه برای هر ( ccTLD (country code top-level domain که در خارج از کشور صورت گرفته است (نامعلوم)

علاوه بر ناقص بودن آمارها، موضوع مهمی که جلب توجه می‌کند، استفاده از شاخص‌های نامناسب برای پایش وضعیت محتوای الکترونیکی کشور است که به طور قطع پوشش دهنده ابعاد مختلف مساله محتوای کشور به خصوص در حوزه وب نبوده و نیاز به بازنگری دارد. در زمینه تولید و نشر محتوای الکترونیکی علی‌رغم وجود الزامات قانونی، شاهد تغییر محسوسی در زمینه تولید، نشر و استفاده از محتوای الکترونیکی فارسی نیستیم. در حالت کلی مشکلات مربوط به محتوای الکترونیکی فارسی را می‌توان در سه دسته مختلف قرار داد که در ادامه مهمترین آنها، تشریح گردیده‌اند.

### ۷-۱- مشکلاتی که بر عدم وجود محتوای الکترونیکی دلالت دارند

- ناکافی بودن قوانین لازم برای نشر محتوای الکترونیکی
- مناسب نبودن کسب و کارهای موجود در زمینه تولید و ارائه محتوای الکترونیک **♦**
- عدم وجود بستر لازم برای ارائه محتوای الکترونیکی به صورت یکپارچه **♦**
- معدود بودن تولید کنندگان و جمع‌آوری کنندگان داده (Data Aggregator)
- ناکافی بودن مهارت‌های لازم برای تهیه محتوای الکترونیکی
- عدم دسترسی به فناوری و الزامات تهیه محتوای الکترونیکی **♦**
- نبود مرجع رسمی شناخته شده برای ارزش گذاری محتوای الکترونیکی

### ۷-۲- مشکلاتی که بر دسترسی به محتوای الکترونیکی دلالت دارند

- ملاحظات امنیتی
- مشکلات مالی و کسب و کار
- ضوابط دست‌یابگیر دولتی
- موانع بین المللی و تحریم

<sup>۱</sup> درگاه پایش جامعه اطلاعاتی، <http://mis.iran.ir/fa/indicators/usage-of-ict/e-content>