مجله ماشین‌بینایی و پردازش‌تصویر
سال پنجم، شماره دوم، زمستان ۱۳۹۷

شبکه‌ی تصویری تصویرنت (تجربیات، چالش‌ها و ابزارهای بومی‌سازی شبکه‌ی تصویری ImageNet)

فرزاد زرگری[۱]، فرزانه رحمانی[۲]، مژگان فرهودی[۳]، محمدحسین زابل زاده[۴]، زینب پرکار[۵]، احسان قاسمی[۶]

چکیده

پایگاه‌های دادگان تصویری بزرگ به عنوان محتوای آموزشی در کاربردهای هوش بصری و همچنین الگوریتم های یادگیری عمیق مورد استفاده قرار می‌گیرند. شبکه‌ی تصویریِ تصویرنت، پایگاه داده‌ای سلسله‌مراتبی از تصاویر و مطابق با فرهنگ ایرانی و اسلامی می‌باشد که با فراهم آوردن حدود ۸ میلیون تصویر برای بیش از ۳۰هزار واژه تهیه شده است. سلسله مراتب ترادف‌های تصویرنت مبتنی بر سلسله مراتب ترادف‌های ImageNet بوده که با استفاده از معادل‌یابی خودکار در زبان فارسی معادل‌یابی شده‌اند. در این شبکه تصویری ۷۸۹۰۷۴۵ تصویر (برای ۳۲۲۹۵ ترادف فارسی شده) با بهره‌گیری از بارگیری خودکار لینک‌های اصلی ارائه شده در ImageNet فراهم گردیده‌اند. سپس ۷۱۸۷۳ تصویر با روش جمع‌سپاری برای بیش از ۱۰۰۰ ترادف مخصوص فرهنگ ایرانی و اسلامی تهیه و به تصویرنت اضافه شده‌اند. هدف از این مقاله انتقال تجربیات، چالش‌های تولید و ابزارهای مورد نیاز حاصل از تولید پایگاه‌داده‌ی سلسله‌مراتبی تصویرنت و همچنین تشریح فرآیند جمع‌سپاری صورت گرفته جهت تهیه تصاویر است.

کلیدواژه‌ها

پایگاه داده‌ی سلسله‌مراتبی تصاویر، شبکه‌ی تصویری فارسی، جمع‌سپاری

۱ مقدمه

به منظور درک صحیح و تشخیص درست محیط پیرامون توسط ماشین ها، علم بینایی ماشین مجاب به استفاده از یک پایگاه داده‌ی سلسله مراتبی ترادف‌ها به نام شبکه‌ی تصویری می‌شود به گونه ای که قدرت بینایی و پردازش عامل های نرم افزاری هوشمند تولید شده به قدرت تشخیص و درک انسان از محیط پیرامون نزدیک تر گردد. شبکه‌ی تصویری یک پایگاه داده‌ی عظیم از تصاویر است که در یک ساختار سلسله مراتبی طبقه بندی شده‌اند[۱]. این ساختار سلسله مراتبی یک ساختار سلسله مراتبی از ترادف‌ها[1] است، که مبتنی بر یک شبکه‌ی کلمات می باشد که برای ساخت شبکه‌ی تصویری به هر کدام از ترادف‌ها از مقوله‌ی اسم که قابل تصویر شدن باشد، تعداد زیادی تصویر نسبت داده می‌شود.

این مقاله در اسفندماه ۱۳۹۵ دریافت، در مهرماه ۱۳۹۶ بازنگری و در آذرماه پذیرفته شد.

این پژوهش با پشتیبانی مالی پژوهشگاه ارتباطات و فناوری اطلاعات بر اساس قرارداد شماره ۷۰۵۷/ص/۵۰۰/پ انجام شده است.

[۱] عضو هیات علمی، [۲] دانشجوی دکتری، [۳] عضو هیات علمی، [۴] همکار پروژه، [۵] همکار پروژه و [۶] همکار پروژه پژوهشگاه ارتباطات و فناوری اطلاعات

رایانامه:
(zargari, rahmani, farhoodi, zabolzadeh, z.porkar, e.ghasemi) @itrc.ac.ir

نویسنده مسئول: فرزاد زرگری

[1] Synset

برای شبکه‌ی تصویری کاربردهای متفاوتی از بالابردن قدرت تشخیص تصاویر در عامل های نرم افزاری مانند تشخیص اشیا در دوربین های امنیتی و رانندگی خودکار تا کاربرد در جویشگرهای تصویری مدنظر قرار می گیرد. در جویشگرهای تصویری، تشخیص اشیا و استنباط منطقی و توصیف حالات از تصویر در زمان پرسش، تشخیص اشیا و استنباط منطقی و توصیف حالات از تصویر در زمان خزش و نمایه گذاری، گسترش پرس و جوی تصویری و نگاشت ترادف به پرسش در جستجوی تصویر و کاهش فاصله‌ی معنایی ترادف و پرسش از جمله زمینه‌هایی است که شبکه تصویری می تواند مفید واقع شود.

ImageNet [۲] ، شبکه‌ی تصویری ساخته شده در دانشگاه استنفورد می‌باشد که مبتنی بر ترادف‌های اسم قابل تصویرشدن موجود در سلسله‌مراتب ترادف‌های وردنت انگلیسی [۳] ساخته شده است.

بومی سازی تکنولوژی های موجود در دنیا همواره یکی از مهم ترین دغدغه های جوامع در حال توسعه می باشد. بومی سازی پروژه‌ها و تکنولوژی های موجود در جهان علاوه بر کاستن وابستگی به جهان و خودکفایی بیشتر ضرورت های بسیاری دارد. شبکه‌ی تصویری تصویرنت[1] با هدف بومی سازی شبکه‌ی تصویری ImageNet و اضافه نمودن ترادف‌ها و تصاویر ایرانی و اسلامی به آن ایجاد شده است[۶،۷]. نسخه‌ی اولیه‌ی تصویر نت در مجموع شامل حدود ۸ میلیون تصویر می‌باشد که در قالب بیش از ۳۳هزار ترادف گنجانده شده‌اند و سلسله مراتب تصویرنت شامل بیش از ۱۰۰۰ ترادف کاملاً مختص فرهنگ ایرانی و اسلامی می باشد که تصاویر مربوط به این ترادف‌ها نیز از طریق جمع‌سپاری برچسب‌گذاری شده و به مجموعه‌ی تصاویر تصویرنت اضافه شده‌اند. شکل ۱ نمایی کوچک از شبکه تصویری تصویرنت را نمایش می دهد. در این شکل سه مسیر از ریشه تابرگ‌ها در سلسله مراتب نمایش داده شده است و برای هر برگ تصویری نمونه ارائه شده است.

در این مقاله ضمن ارائه‌ی دستاوردهای شبکه‌ی تصویری تصویرنت و جمع سپاری انجام شده، چالش‌ها و تجربیات ایجاد این پروژه برای دیگر پژوهشگران که در زمینه‌های مشابه فعالیت می‌کنند، تشریح می‌شود. ادامه‌ی مقاله به صورت زیر ساختاربندی شده است. در فصل دوم شبکه‌ی تصویری Imagenet به همراه شبکه‌ی کلمات وردنت معرفی می شوند. در فصل سوم چالش‌ها و تجارب ساخت شبکه‌ی تصویری فارسی تصویرنت ارائه می‌گردد. فصل چهارم فرآیند افزودن ترادف‌های ایرانی و اسلامی و جمع‌سپاری تصاویر مربوط به آن‌ها را به تشریح بیان می‌کند. فصل پنجم آماری کلی از تصویرنت را فراهم نموده است و در ادامه فصل ششم جمع‌بندی و آینده‌ی تصویرنت را شرح می‌دهد.

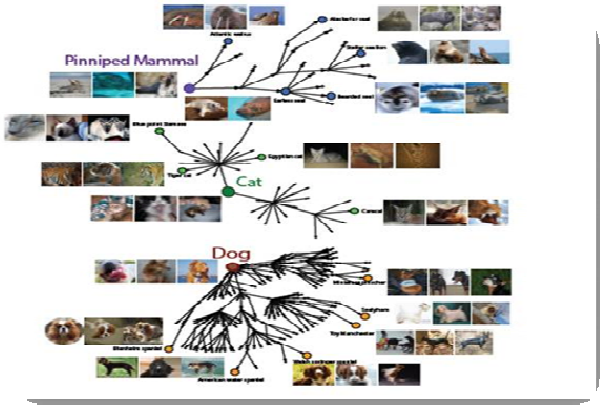
۲ معرفی شبکه تصویری ImageNet و شبکه کلمات وردنت

شبکه‌ی تصویری Imagenet به طور گسترده در تحقیقات پردازش تصویر مانند آموزش شبکه‌های عصبی، یادگیری عمیق، برچسب‌زنی خودکار تصاویر و همچنین تحقیقات گوگل [۴و۵] مورد استفاده قرار گرفته است. ایده‌ی اصلی ImageNet از آن جا ناشی شد که بررسی یک کودک در سه سالگی نشان داد که او بدون هیچ آموزش و تخصصی می تواند تصاویر را به درستی ببیند و تشخیص دهد. نکته‌ی حایز اهمیت این است که در سنین ابتدایی این مسائل به کودکان آموزش داده نمی شود و تشخیص تنها با تکرار دیدن تصاویر صورت می‌گیرد. این مشاهده به ساخت شبکه ای تصویری برای اشیا انجامید.

شکل ۱- نمایی از سلسله مراتب تصویرنت

در صورتیکه به ازای هر شی تعداد قابل توجهی تصویر با کیفیت بالا و ابعاد مناسب جمع آوری شود، تنها کافی است این تصاویر به عنوان ورودی آموزشی به عامل نرم افزاری داده شود. هدف پروژه ImageNetجمع‌آوری تعداد زیادی تصویر در ابعاد و کیفیت بالا به صورت یک ساختار شبکه‌ای منسجم بوده است تا به طور متوسط برای هر ترادف ۱۰۰۰ تصویر وجود داشته باشد. این شبکه‌ی تصویری در حال حاضر تعداد ۱۴ میلیون تصویر را برچسب‌گذاری نموده است و در قالب ۳۲هزار گره یک گراف همبند بدون دور که هر کدام یک ترادف اسم از وردنت به عنوان برچسب این تصاویر هستند، طبقه‌بندی نموده است. میانگین تعداد تصاویر برای هر ترادف Imagenet برابر ۴۳۹ عدد می‌باشد و ۲۱۸۴۱ گره آن دارای تصویر می باشند. شکل ۱ بخش کوچکی از شبکه‌ی تصویری ImageNet شامل زیر کلاس های Cat و Dog را به همراه تصاویر نمونه‌ی آن نمایش می دهد.

این شبکه‌ی تصویری در حال حاضر دارای ۸ کلاس اصلی از تصاویر به همراه یک کلاس از تصاویر برچسب گذاری شده که در این ۸ کلاس نمی گنجند، می باشد. این کلاس‌ها به همراه تعداد زیر کلاس‌های‌شان در جدول ۱ ارائه شده‌اند.



شکل ۲- نمایی کوچک از ساختار سلسله مراتبی و تصاویر ImageNet

جدول۱- کلاس‌های اصلی تصاویر در ImageNet به همراه کلاس متفرقه

نام کلاس	معادل فارسی	تعداد زیر کلاس
Plants	گیاهان	۴۴۸۶
Geological formation	سازند زمین سناسی	۱۷۵
Natural Objects	اشیای طبیعی	۱۱۱۲
Sport, Athletics	ورزش	۱۷۶
Artifacts	مصنوعات	۱۰۵۰۴
Fungus	قارچ	۳۰۸
Person	افراد	۶۹۷۸
Animal	حیوانات	۳۹۹۸
Mics	متفرقه	۲۰۴۰۰

وردنت انگلیسی یا شبکه واژگانی پرینستون نخستین بار توسط میلر و همکاران[1] در آزمایشگاه علوم شناختی دانشگاه پرینستون بر اساس واژگان ذهنی و در حوزه ی پژوهش‌های روانشناسی زبان، طراحی و ایجاد شد. وردنت در واقع شبکه‌ای معنایی از بیش از یکصد هزار ترادفی است که بوسیله روابط معنایی به هم مرتبط هستند. هر ترادف نشان‌دهنده‌ی مجموعه‌ای انتزاعی از عناصری می‌باشد که بر اساس مختصه‌های مشترکشان یک گروه را تشکیل می‌دهند. وردنت انگلیسی یک پایگاه داده لغوی بزرگ از لغات انگلیسی است. این بانک اطلاعاتی، اسم‌ها، فعل‌ها، صفت‌ها و قیدها را به مجموعه‌ای از لغات مترادف به نام ترادف دسته‌بندی می‌نماید. ساختار وردنت انگلیسی، از آن یک ابزار مفید جهت زبان‌شناسی رایانشی و پردازش زبان طبیعی بوجود آورده است. تعداد اسامی موجود در وردنت انگلیسی ۸۲۱۱۵ مورد می‌باشد.

۳ چالش‌ها و تجارب ساخت شبکه تصویری فارسی(تصویرنت)

شبکه‌ی تصویری تصویرنت از سلسله‌مراتب ترادف‌های ImageNet استفاده می‌کند که مبتنی بر وردنت بوده و به زبان انگلیسی می‌باشد. سلسله مراتب استفاده شده در ImageNet بسیار کامل و جامع بوده و برای کاربردهای پردازش تصویر مناسب است. در جهت بومی‌سازی این سلسله‌مراتب ترادف نیاز است تا ترادف‌های آن به زبان فارسی معادل‌سازی شوند. همچنین تصاویر مربوط به ImageNet به صورت خودکار بارگیری شده و با تشکیل پایگاه داده‌ی گرافی Neo4j، این ترادف‌ها و تصاویر یک شبکه‌ی تصویری با زبان فارسی را تشکیل دهند. در نسخه‌ی اول تصویرنت، فارسی بودن سلسله مراتب ترادفی برای بیش از ۹۳% واژگان محقق شده است و حدود ۸ میلیون تصویر بارگیری شده‌اند. در ادامه، نخست نحوه‌ی ساخت سلسله مراتب ترادف‌های تصویرنت به زبان فارسی شرح داده می‌شود. سپس به چالش‌های بارگیری خودکار تصاویر پرداخته می‌شود و جزئیات ساخت پایگاه داده‌ی Neo4j بیان می‌شود.

۳-۱ ساخت سلسله مراتب ترادف‌ها به زبان فارسی

برای ساخت شبکه‌ی تصویری، ابتدا یک سلسله مراتب جامع از ترادف‌ها مورد نیاز است. با توجه به کامل و جامع بودن سلسله‌مراتب وردنت، تصویرنت نیز مشابه ImageNet از این سلسله مراتب شامل ترادف‌های انگلیسی استفاده نموده است. در قدم بعدی روشی برای معادل‌سازی ترادف‌های شبکه‌ی ترادف‌های وردنت به زبان فارسی مدنظر قرار گرفت[۷]. یک روش ساده برای این کار استفاده از یک لغت‌نامه و ترجمه خودکار عبارات است.

[1] George A. Miller et. al.

اما گاهی عبارات و کلمات موجود در ترادف‌ها لزوماً معانی لغوی مرتبط با خود را تداعی نمی‌کنند و لازم است که از معادلی که به صورت مصطلح برای آن کلمه یا عبارت در اختیار است استفاده شود. باید توجه داشت که این کار باید در مورد حجم بالایی از ترادف‌ها و نیز به صورت خودکار انجام شود. لذا کلمات و عبارات موجود در ترادف‌ها شکسته شده و پس از انجام پیش‌پردازش بر روی این کلمات و عبارات و حذف مواردی مانند فاصله و نیم فاصله مجموعه‌ای از کلمات حاصل شد.

برای معادل‌یابی خودکار این کلمات، منابعی شامل ترگمان[۸] که اولین سامانه ترجمه ماشینی آماری دوزبانه در ایران است به همراه ترجمه‌ی گوگل[۹]، ویکی‌پدیا[۱۰] و شبکه‌ی کلمات فارس‌نت [۱۱] مد نظر قرار گرفته و برای تمامی کلمات با بکارگیری رابط برنامه نویسی کاربردی[1] آن‌ها معادل‌سازی خودکار انجام شد. سپس به طور نمونه و تصادفی نتایج ترجمه‌ی خودکار حاصل از این منابع برای ۹۸ ترادف توسط خبره انسانی مقایسه و بررسی شد. برای معادلی که هر مترجم برای هر ترادف در اختیار قرار داده است، امتیازی در نظر گرفته شد. این امتیازات در جدول ۲ شرح داده شده‌اند.

جدول۲ امتیازات در نظر گرفته شده برای نتایج مترجم ها

امتیاز	شرح امتیاز
۱	مرجع موردنظر نتوانسته ترجمه‌ای را در اختیار قراردهد و هیچ خروجی‌ای نداشته است.
۲	مرجع موردنظر ترجمه داده است ولی ترجمه ارائه شده ارتباطی با معنی کلمه مورد نظر ندارد.
۳	ترجمه خوب
۴	ترجمه خیلی خوب
۵	ترجمه عالی

در جدول ۲ امتیازات ۳ ، ۴و ۵ به این دلیل در نظرگرفته شد که بین نتایج درست چند مرجع رتبه بندی انجام داده و برتری مترجم‌ها را نسبت به هم بهتر نمایش داد. تعداد خروجی‌های هر مترجم بر حسب امتیازات در نظر گرفته شده به صورت جدول ۳می‌باشد.

جدول۳ امتیاز(تعداد کلمه) حاصل برای هر مرجع

امتیاز مترجم	۱	۲	۳	۴	۵
ویکی پدیا	۴۰	۱۳	۱۸	۸	۱۹
فارس نت	۹۳	.	۱	۰	۴
ترگمان	۷	۱۷	۱۴	۱۲	۴۷
گوگل	۰	۳۱	۳۲	۱۰	۲۵

بنا بر جدول ۳ نتایج بدست آمده نشان داد که فارس‌نت با این‌که برای تعداد کمی از کلمات معادل داشته است،ولی این معادل‌ها کاملاً درست بوده‌اند. همچنین مشخص شد که ویکی‌پدیا در کلماتی که دارای ریشه علمی بوده یا اسم شخص است (به طور کلی اسم خاص) بهتر عمل کرده است. در خروجی‌هایی که از ترگمان بدست آمده، مشخص شد که ترجمه‌هایی که دارای عبارتی قرارگرفته در داخل پرانتز باشند، با احتمال خوبی درست هستند. گوگل نیز برای همه ترجمه‌ها خروجی داده که البته در بسیاری از موارد همان عبارت انگلیسی را در خروجی نشان داده است. در نتیجه بر مبنای تمامی مشاهدات صورت گرفته یک الگوریتم نهایی بر روی تمامی ترادف‌ها اجرا شد.

شروع

قدم اول: ابتدا فارس‌نت بررسی شود اگر ترجمه داشت همان به عنوان ترجمه انتخاب می‌شود و پایان معادل یابی ترادف.

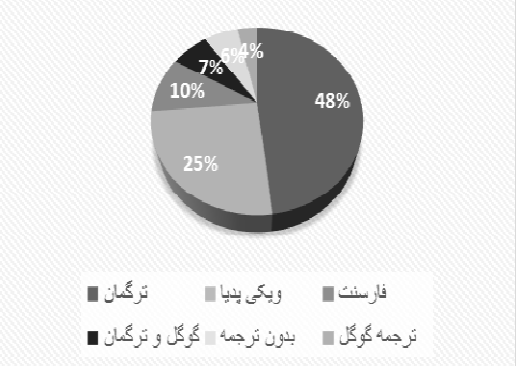
قدم دوم: اگر ترگمان یک کلمه را به عنوان ترجمه نتیجه داده باشد، انتخاب می‌شود و پایان.

قدم سوم: اگر ویکی‌پدیا ترجمه داشت، انتخاب می‌شود و پایان.

قدم چهارم: اگر هیچکدام از حالات قبل اتفاق نیفتاده باشد، از ترجمه‌های گوگل و ترگمان ترجمه‌ای انتخاب می‌شود که دارای فراوانی بیشتری باشد.

پایان

در نهایت بیش از ۹۳% از ترادف‌ها دارای معادل فارسی می‌باشند[۷]. برای تعدادی از ترادف‌ها هیچ معادلی در زبان فارسی تعریف نشده است برای مثال "Gyromitra" که نوعی قارچ است و یا گیاه "myrmecophyte "نام معادل فارسی ندارد و لذا این برچسب‌ها به صورت انگلیسی در تصویرنت درج شده‌اند. شکل ۳ تعداد و درصد ترادف‌های معادل‌یابی شده را به تفکیک مرجع نمایش می‌دهد.با توجه به اینکه هر کدام از ترادف‌ها در وردنت یک عبارت توصیفی دارند که واژه مورد نظر را شرح می‌دهد، نیاز است این توصیف‌ها در تصویرنت به صورت فارسی معادل‌سازی شوند. برای این منظور از ترجمه‌ی گوگل و فارس‌نت برای معادل‌سازی توصیف وردنت استفاده شده است.



شکل۳ جزئیات مراجع معادل یابی سلسله مراتب تصویرنت

[1] API

۳-۲ بارگیری خودکار تصاویر

برای ایجاد دادگان تصویری تصویرنت مشابه آنچه در شبکه‌ی تصویری ImageNet ارائه شده است و به صورتی که مستقل از سیاست‌های این شبکه‌ی تصویری باشد، مجموعه تصاویر موجود در آن تا حد ممکن بارگیری شده‌اند. طبق قوانینی که در خصوص بارگیری این مجموعه‌داده در نظر گرفته شده، امکان بارگیری از ImageNet تنها برای مقاصد تحقیقاتی امکان‌پذیر است و لذا با توجه به اینکه ممکن است در آینده از این مجموعه‌داده برای مقاصدی تجاری در ایران استفاده شود، تصمیم گرفته شد که تصاویر این پایگاه به صورت یک‌به‌یک از منبع اصلی آن‌ها تهیه شوند. در نتیجه، از شبکه‌ی تصویری ImageNet منبع اصلی برای این تصاویر را که به صورت یک پرونده شامل مجموعه پیوندهای منابع اصلی تصاویر بود تهیه شد، که با استفاده از آن‌ها می‌توان تصاویر را از مرجع اصلی بدست آورده و بارگیری نمود.

برای بارگیری این تصاویر که تا زمان حاضر در حدود ۱۴ میلیون تصویر هستند، نیاز به ابزاری است که به صورت خودکار و با بررسی هر یک از پیوندها، تصویر مربوط به آن را بارگیری نماید. البته برای انجام این کار چالش‌هایی وجود دارد که در ادامه به آن‌ها اشاره شده و راه‌حل‌های ممکن و قابل پیاده‌سازی برای آنها مورد ارزیابی قرار گرفته‌اند.

برخی از این چالش‌ها شامل موارد زیر می‌باشد:

- تعداد زیاد پیوندها
- یکسان نبودن سرورهای حاوی تصاویر و استاندارد نبودن مدت زمان پاسخ سرورها
- عدم امکان دسترسی به تعداد زیادی از پیوندها به دلیل محدودیت دسترسی در ایران
- محدودیت پهنای باند و سرعت ارائه شده توسط شبکه اینترنت

برای رویارویی با این چالش‌ها راهکارهایی ارائه شد که بتوان حداکثر تصاویر ممکن را در کوتاهترین زمان بارگیری نمود[۷]. تعدادی از این راهکارها به شرح زیر می‌باشند:

- بارگیری دسته‌ای از پیوندها به طور همزمان
- بهبود استفاده از پهنای باند موجود
- ارتقاء عملکرد بارگیری پیوندها به صورت تک-حلقه و براساس رخدادها
- بارگیری مجموعه پیوندهای هر دسته به صورت غیرهمزمان

از میان پیوندهای ارائه شده توسط ImageNet تعداد زیادی مربوط به پایگاه flickr.com هستند که دسترسی به آنها به دلیل محدودیت دسترسی این سایت در ایران ممکن نیست. لذا لینک‌ها به دو دسته flickr و غیر flickr تقسیم شدند. حدوداً نیمی از پیوندها مربوط به تصاویر پایگاه flickr و مابقی از پایگاه‌های متفرقه دیگر هستند. مجموعه پیوندهای غیر flickr به ۵ قسمت مساوی تقسیم‌بندی شده و در ۵ سیستم مجزا به طور همزمان وارد پروسه بارگیری شدند.

برای انجام بارگیری به دلیل تعداد بالای پیوندها در هر قسمت، استفاده از روش بارگیری یک‌به‌یک چندان به صرفه نبوده و پهنای باند موجود را کاملا مورد استفاده قرار نمی‌دهد، لذا باید تعدادی از پیوندها را به طور همزمان بارگیری نمود. برای این منظور از یک روش برنامه‌نویسی غیر همگام[1] استفاده شده است.

چالش دیگر مدت زمان غیر استاندارد پاسخ سرورها است. برای حل این مسئله از یک زمان‌سنج در ارتباط به سرور مربوطه استفاده شد که در میان همه‌ی پیوندها دارای زمان یکسان باشد. اما در این حالت ممکن است یک پیوند سالم و حاوی تصویر بوده و برنامه مشغول به بارگیری تصویر مربوط به آن باشد، که به دلیل حجم بالای فایل و یا کند بودن ارتباط میان برنامه و سرور به مدت زمان بیشتری از زمان تعیین شده در زمان‌سنج نیاز داشته باشد. راه حلی که برای این مشکل ارائه شد، استفاده از یک برنامه بر اساس برنامه‌نویسی مبتنی بر رخداد است. این برنامه تعداد مشخص شده‌ای از فایل‌ها را از سرور مربوطه مورد درخواست قرار می‌دهد و برای پاسخ حاصل مدت زمانی را منتظر می‌ماند. در صورتی که پاسخی از سرور دریافت شود که حاکی از وجود فایل تصویری است، این فایل دانلود شده و در غیر این صورت از آن گذر خواهد شد. در هر صورت اطلاعاتی مربوط به فرآیند بارگیری در فایلی جداگانه ثبت می‌گردد. اطلاعات ثبت شده شامل پاسخ دریافت شده از سرور، قالب پاسخ دریافت شده و حجم آن و نیز دیگر اطلاعات مربوط به تعداد و لیستی از پیوندها می‌باشد.

پس از اجرا شدن برنامه‌های پیاده‌سازی شده بر روی سیستم‌ها و بررسی کارایی آنها، پارامترهایی مانند میزان تاخیر در بین هر مجموعه از دانلودها، مدت زمان تاخیر برای زمان‌سنج مربوط به بارگیری هر پیوند و نیز تعداد ارتباطات همزمان به صورت خاص برای هر سیستم بررسی و به مرور زمان ارتقاء یافتند.

برای بارگیری تصاویر پایگاه flicker نیز از روش به کارگرفته شده در مورد تصاویر غیر flicker استفاده شد، با این تفاوت که با در اختیار گرفتن سرورهای مجازی بدون محدودیت دسترسی، بارگیری تصاویر flicker برروی این سرورها انجام شده و سپس این تصاویر به سرورهای داخلی انتقال داده شدند.

۳-۳ ساخت پایگاه داده

شبکه‌ی تصویری یک گراف جهت‌دار همبند است که هر یک از ترادف‌ها به عنوان یک گره و ارتباط آن‌ها با تعریف یک یال مشخص می‌شود. این ارتباط از نوع رابطه‌ی پدر-فرزندی است. بهترین ابزار برای ذخیره‌سازی اطلاعات گراف، پایگاه داده‌هایی از نوع پایگاه داده گرافی است که یکی از بالغ‌ترین محصولات متن باز به نام Neo4j برای ذخیره‌سازی اطلاعات گره‌های تصویرنت مورد

[1] Asynchronous

استفاده قرار گرفته است[۷]. در پایگاه داده‌های Neo4j هر موجودیت معادل با یک گره است. هر گره ممکن است با سایر گره‌ها ارتباط داشته باشد، که به آن رابطه گفته می‌شود. هر گره یا رابطه دارای تعدادی خصیصه است که صفت‌های مربوط به آن‌را تشکیل می‌دهند. هر گره ممکن است تعدادی برچسب بپذیرد که معرف جنس و نوع آن گره است. در پایگاه داده‌ی تصویرنت هر یک از گره‌ها دارای خصیصه‌هایی هستند که در جدول ۵ به همراه توضیحی از آن‌ها لیست شده‌اند. لازم به ذکر است که تعدادی از این خصیصه‌ها می‌توانند در برخی نودها مقدار نداشته باشند.

در خصیصه‌های مربوط به هر گره در جدول ۴ گزینه‌ای با نام visible در نظر گرفته شده است که امکان نمایش یا عدم نمایش آن گره را در ساختار گراف فراهم می‌سازد. از این طریق در صورتی که ترادفی با شئونات ایرانی و اسلامی مغایرت داشته باشد، می‌توان آن را در ساختار گراف مخفی نمود. در واقع تصاویر ImageNet حذف نمی شوند و فقط آن هایی که با شئونات ایرانی و اسلامی مطابقت ندارند به صورت invisible درآمده و تنها با مجوز های خاص در اختیار کاربران قرار می گیرند. برای مثال ImageNet یک زیرشاخه‌ی بزرگ از انواع نوشیدنی‌های الکلی ارائه کرده است. نظر بر است که این زیرشاخه به صورت invisible باشد اما در بحث پردازش تصاویر گمرک و جلوگیری از قاچاق این نوشیدنی ها، وجود تصاویر آن‌ها ضروری است. در حال حاضر این زیر شاخه هنوز invisible نشده است. اما تعدادی از ترادف‌ها مربوط به موضوعات جنسی invisible شده اند. این نکته در واقع گامی به سوی بومی سازی شبکه‌ی تصویری برمی دارد.

جدول۴ خصیصه‌های تعریف شده در هر گره گراف

نام خصیصه	شرح
tnid	شناسه‌ی تصویرنت
wnid	شناسه‌ی وردنت
fnid	شناسه‌ی فارسنت
wnwords	کلمات وردنت
wngloss	توصیف وردنت
fnwords	کلمات فارسنت
fngloss	توصیف فارسنت
tnwords	کلمات تصویرنت
tngloss	توصیف تصویرنت
trnswords	کلمات ترجمه شده
trnsgloss	توصیف ترجمه شده
trnssrc	منبع ترجمه
visible	قابل دیده شدن یا غیر قابل دیده شدن در ساختار گراف

۴ افزودن ترادف‌های ایرانی واسلامی و جمع‌سپاری تصاویر مربوط به آن‌ها

در فرهنگ ایرانی و اسلامی بسیاری از ترادف‌ها و اشیا وجود دارند که در سلسله مراتب ImageNet تعریف نشده‌اند. برای تطبیق هر چه بیشتر شبکه‌ی تصویری تصویرنت با فرهنگ ایرانی اسلامی ۱۰۰۰ ترادف که مختص فرهنگ ایرانی اسلامی می‌باشند، انتخاب شده و به آن اضافه شده‌اند. استراتژی تهیه‌ی تصاویر برای ترادف‌ها ایرانی و اسلامی استفاده از جمع‌سپاری برای برچسب‌گذاری تصاویر به کمک تعداد زیادی از کاربران می‌باشد[۷].

برای انجام جمع‌سپاری سامانه‌ی تحت وب جمع‌سپاری با روال‌های مشخص و پیش‌نیازهایی نظیر دادگان جمع‌سپاری مورد نیاز می‌باشد. از این رو ابتدا سامانه‌ی جمع‌سپاری، توسعه داده شده و روال‌های برچسب‌گذاری تصاویر در آن مشخص شد. سپس پیش‌نیازهای بانجام جمع‌سپاری فراهم گردید. این پیش‌نیازها، مواردی نظیر تهیه‌ی داده‌ی مربوط به جمع‌سپاری اعم از داده‌های تصویری و اطلاعات مورد نیاز برای برچسب‌گذاری تصاویر و همچنین فراخوان و دعوت کاربران برای شرکت در این فعالیت را در برمی‌گیرد. سپس فرآیند جمع‌سپاری با دو فعالیت اصلی آموزش و ارزیابی برچسب‌گذاری تصاویر و فعالیت اصلی برچسب‌گذاری تصاویر انجام شده و پس از تحلیل نتایج بیش از ۷۰ هزار تصویر برچسب‌گذاری شده به تصویرنت اضافه گردید. شکل ۴ نمای کلی فعالیت‌هایی که لازم است در فرآیند جمع‌سپاری انجام شود نشان می‌دهد. در ادامه چگونگی انتخاب ترادف‌های ایرانی‌واسلامی به همراه جمع آوری تصاویری که باید جمع‌سپاری برچسب‌زنی برای آن‌ها انجام شود، شرح داده می‌شود. سپس پیش‌نیازهایی که برای جمع‌سپاری فراهم شده است، بیان شده و روال‌های جمع‌سپاری تشریح می‌شود.

۴-۱ انتخاب ترادف‌های ایرانی و اسلامی برای افزودن به تصویرنت

برای ترادف‌های جدید قابل اضافه شدن به شبکه‌ی تصویری بررسی دو دسته ترادف‌ها مد نظر قرار گرفت.

دسته‌ی اول: ترادف‌های اسم قابل تصویرشدن از فارسنت که مختص فرهنگ ایرانی و اسلامی می‌باشند و معادلی در وردنت ندارند.

دسته‌ی دوم: ترادف‌های اسم قابل تصویرشدن از مراجع دیگری انتخاب شوند. برای فراهم کردن ترادف‌های جدید، دانشنامه‌ی ایران[۱۲] و دائره‌المعارف بزرگ اسلامی [۱۳] به همراه مراجعی مانند ویکی پدیای فارسی [۱۰] مورد استفاده قرار گرفتند. زیرا این مراجع، ترادف‌های ایرانی و اسلامی را به صورت یکجا در اختیار قرار می‌دهند و مرجعی مانند ویکی‌پدیا برای ترادف‌ها دسته‌بندی اولیه‌ای را نیز فراهم آورده است. پس از تشخیص اسامی قابل تصویر شدن، بررسی‌هایی انجام شد تا ترادف‌های تکراری در سلسله‌مراتب درج نشود. علاوه بر ترادف‌های ذکر شده، ترادف‌هایی از بناهای تاریخی و باغ‌های ملی ایرانی و شخصیت‌های کارتونی و عروسکی نیز برای اضافه شدن به سلسله‌مراتب در لیست ترادف‌های ایرانی و اسلامی قرار گرفتند.

هر یک از ترادف‌های موجود در سلسله‌مراتب به همراه یک توصیف کوتاه از آن ترادف در پایگاه داده‌ی گرافی ذخیره می‌شود تا تصویرنت در هنگام نمایش دادن نتایج پرس و جو و یا کاربردهای دیگر بتواند ترادف را به صورت کاملتر معرفی نماید. در نتیجه برای این ترادف‌های جدید نیز عبارتی توصیفی تهیه گردید.
پس از ایجاد گره‌های جدید برای هریک از ترادف‌های ایرانی واسلامی، این ترادف‌های جدید در کلاس[1] هایی دسته بندی شدند که در محل اضافه شدن آن‌ها به گراف مفید واقع شد.

با بررسی ترادف‌های سلسله‌مراتب توسط کاربر انسانی، مکان مناسب درج گره‌های جدید تشخیص داده شد. جدول ۵ کلاس‌های ترادف‌های جدید را نمایش می دهد.

جدول۵کلاس های اصلی ترادف‌های ایرانی واسلامی

شمار زیرگروه	نام زیر گروه
۱	آیین‌ها و مراسم مذهبی و اسلامی
۲	آیین‌های ایرانی
۳	غذاها و لباس‌های سنتی ایرانی
۴	معماری ایرانی و اسلامی
۵	هنرهای دستی ایرانی
۶	خوشنویسی ایرانی و اسلامی
۷	ورزش‌هاو بازیهای ایرانی و باستانی
۸	پوشش و حجاب ایرانی و اسلامی
۹	بناهای تاریخی و باغ‌های ملی ایرانی
۱۰	شخصیت‌های کارتونی و عروسکی
۱۱	ترادف‌های ایرانی و اسلامی متفرقه

۴-۲ جمع آوری تصاویر مربوط به ترادف‌های ایرانی و اسلامی برای جمع‌سپاری

از نتایج تصویری بازگردانده‌شده توسط جویشگرهای تصویری گوگل [۱۴]، بینگ [۱۵] و جویشگر ایرانی پارسی‌جو [۱۶] به عنوان تصاویر جمع‌سپاری استفاده شد. برای این منظور برای هر ترادف یک عبارت پرس و جو تعیین شد که بتواند به خوبی مشخص کننده‌ی ترادف مورد نظر باشد و مرتبط ترین تصاویر از جویشگرها دریافت شود. به عنوان مثال از عبارت "مُهر نماز" برای جستجوی مُهر استفاده شد. این عبارت‌های پرس‌وجو به صورت خودکار به سه جویشگر مطرح شده داده شدند و ۲۰۰ نتیجه‌ی تصویری ابتدایی حاصل از آن‌ها به صورت url در فایل‌های متنی ذخیره شدند. این urlها به صورت خودکار دانلود شدند و تصاویر دانلود شده نیز با شناسه‌ی url ذخیره سازی شدند. در این مرحله ۲۲۸,۵۲۶ تصویر بارگیری شد.

۴-۳ تهیه‌ی داده‌ی مورد نیاز جمع‌سپاری برای دو بخش آموزش و ارزیابی و برچسب‌گذاری تصاویر

اطلاعاتی مربوط به تعدادی از ترادف‌های انتخاب شده برای این فعالیت مورد نیاز است. این اطلاعات شامل موارد زیر می‌باشد:

- یک توصیف. این توصیف از مراجعی مانند ویکی‌پدیا تهیه شده است.
- عبارتی برای توضیح بیشتر که شکل ظاهری آن ترادف را بیشتر مشخص می‌سازد.
- یک تصویر نمونه از آن ترادف
- یک سوال در مورد توصیف و توضیح آن ترادف.
- تعدادی تصویر شامل تصاویر مرتبط و تصاویر نامرتبط با ترادف مورد نظر.

دسته‌ی دوم داده‌های مورد نیاز برای انجام برچسب‌گذاری تصاویر می‌باشند که برای ۱۰۰۰ ترادف ایرانی واسلامی تهیه شده‌اند:

- یک توصیف،
- عبارتی برای توضیح بیشتر که شکل ظاهری آن ترادف را بیشتر مشخص می‌سازد،
- یک تصویر نمونه از آن ترادف
- تصاویر جمع آوری شده برای آن ترادف

۴-۴ رصد فعالیت کاربران

به منظور رصد فعالیت کاربران در برچسب‌گذاری تصاویر مربوط به یک ترادف و پیشگیری و حذف انتخاب‌های تصادفی کاربران درصورت مشاهده، دقت نسبی هر کاربر در طول انجام فعالیت‌ها محاسبه شده است. در نتیجه بیش از دو درصد از تصاویر مربوط به هر ترادف توسط چند فرد خبره برچسب‌گذاری شد و برای هر مجموعه تصاویر مربوط به یک ترادف تعدادی از آن‌ها که مرتبط و یا نامرتبط از لحاظ ظاهر تصویری با ترادف مورد نظر بودند، مشخص شدند. این تصاویر برچسب‌گذاری شده در بین تصاویر تخصیص داده شده به یک فرد قرار گرفته و بر اساس جوابی که کاربر برای آن‌ها انتخاب کرده و پاسخی که تیم تصویرنت برای این تصاویر در نظر گرفته است، دقت نسبی کاربر به صورت درصد مشخص می‌شود.

۴-۵ فرآیند جمع‌سپاری و کاربران آن

پس از اعلان برای همکاری در فعالیت جمع‌سپاری، ۷۲ نفر برای همکاری اعلام آمادگی کردند. اکثر این افراد فارغ‌التحصیل یا دانشجوی رشته‌های کامپیوتر و فناوری اطلاعات در مقاطع مختلف کارشناسی، کارشناسی ارشد و یا دکتری بوده و در بازه‌ی سنی ۱۸ تا ۴۰ قرار گرفتند. در این مرحله کاربران پس از ثبت نام در سامانه، در یک مرحله آموزش و ارزیابی شرکت داده شده و در نتیجه کسانی که با موفقیت این مرحله را پشت سر گذاشته بودند و تعداد آن‌ها ۶۴ نفر بود، وارد فعالیت اصلی برچسب‌گذاری تصویر شدند. فعالیت جمع‌سپاری در طول مدت ۲ هفته به انجام رسید.

[1] Class

شکل ٤-مراحل مختلف فرآیند جمع‌سپاری

۴-۶ فعالیت «آموزش و ارزیابی کاربران در برچسب‌گذاری تصاویر»

در این فعالیت برای هر کاربر ۵ سوال مربوط به ۵ ترادف متفاوت از میان ترادف‌های که انتخاب شده بود، در نظر گرفته شد. هر سوال از سه بخش آموزش، ارزیابی اولیه و ارزیابی نهایی تشکیل شده که در قسمت آموزش با نمایش یک ترادف اسم به کاربر، توضیحی به همراه اطلاعات بیشتر در مورد کلمه مورد نظر ارائه شده است که به توصیف بیشتر آن می‌پردازد. لازم به ذکر است که اطلاعات بیشتر برای برخی ترادف‌ها که نیاز به توصیف بیشتری در جهت درک ترادف دارند، در اختیار کاربران قرار گرفته است. یک تصویر نمونه نیز از ترادف مورد نظر به کاربر نمایش داده شده است. درقسمت ارزیابی اولیه، یک سوال در مورد ترادف آموزش داده شده به کاربر نمایش داده شده که به صورت چهارگزینه‌ای است. آخرین قسمت برای هر سوال، ارزیابی نهایی است که در آن، تعدادی تصویر به کاربر نشان داده شده که کاربر در صورت تشخیص مرتبط بودن تصویر با کلیک کردن بر روی آن، تصویر مورد نظر را انتخاب می‌کند. در صورت انتخاب اشتباه، برای کاربر امتیاز منفی در نظر گرفته می‌شود و بر مبنای انتخاب تصاویر صحیح و عدم انتخاب تصاویر غلط به کاربر امتیاز داده می‌شود. در صورتی که کاربر حداقل ۷۵ درصد تصاویر را صحیح تشخیص دهد، این سوال برای او به عنوان سوال درست لحاظ می‌شود. کاربرانی که حداقل به سه سوال از ۵ سوال پاسخ صحیح داده باشند موفق شده‌اند از مرحله‌ی آموزش و ارزیابی گذر کرده و مرحله‌ی اصلی فعالیت که برچسب‌گذاری تصاویر است به صورت خودکار در سامانه برایشان فعال می‌گردد.

۴-۷ فعالیت «برچسب‌گذاری تصاویر»

در سامانه برای هر کاربر۱۵۰ کار تعریف شده است. در هر کار یک مرحله‌ی آموزش وجود دارد که مشابه مرحله‌ی آموزش در فعالیت آموزش و ارزیابی می‌باشد. علاوه بر مرحله‌ی آموزش، هر کار حداکثر شامل ۴۰ تصویر است، که به‌صورت صفحات ۴تایی به کاربر نمایش داده می‌شود. کاربر با توجه به توضیح، اطلاعات بیشتر و عکس نمونه، ترادف مورد نظر را درک کرده و به برچسب‌گذاری تصاویر بر مبنای آن ترادف می‌پردازد.

از آن جایی که تصاویر جمع‌سپاری از جویشگرهای مطرح گرفته شده اند، ممکن است تعدادی از تصاویر حاوی موارد غیر اخلاقی باشند و یا تعدادی از آن‌ها به صورت شکسته بارگیری شده باشند یا

کیفیت مناسبی نداشته باشند. برای گزارش این موارد، گزینه‌ی گزارش خرابی تصویر در نظر گرفته شده که با کلیک روی آن، تصویر به عنوان یک تصویر خراب انتخاب و گزارش می‌شود.

۴-۸ تحلیلی بر جمع‌سپاری

در فعالیت جمع‌سپاری انجام شده جمعاً برای ۲۲۸,۵۲۶ تصویر توسط ۶۴ نفر تصمیم‌گیری شد. هر تصویر به سه نفر برای انتخاب

یا عدم انتخاب داده شده و سپس از رأی اکثریت برای تصمیم‌گیری نهایی استفاده شد. در نتیجه مجموعاً ۶۸۵,۵۷۸ فعالیت برای تصمیم‌گیری وجود داشت. با احتساب هر کار حداکثر ۴۰ تصویر و یکتا بودن ترادف مورد نظر در هر کار، ۱۹,۱۰۹ کار در سامانه تعریف شد.

در مجموع با محاسبه‌ی رأی اکثریت تعداد ۶۸۹۴۹ تصویر به صورت مرتبط با ترادف‌های تعیین شده تشخیص داده شدند. و با اضافه نمودن تصاویری که توسط افراد خبره در ابتدا برچسب زنی شده، مجموعاً ۷۱۸۷۳ تصویر از جمع‌سپاری نتیجه شد. تصاویر حاصل از جمع سپاری در فرمت های مختلف تصویری می باشند و دارای کیفیت نسبتاً مناسبی هستند به گونه ای که برچسب مد نظر را به خوبی نمایش دهند. جدول ۶ ویژگی‌های این تصاویر را نمایش می‌دهد.

جدول۶ ویژگی های تصاویر جمع‌سپاری

ویژگی	شرح مقادیر
محدوده رزولوشن برحسب ابعاد به پیکسل	از ۱۰۰×۸۳ تا ۵۷۸۷×۸۱۵۰
فرمت های مختلف تصویری موجود و درصد تصاویر در هر فرمت	JPEG (۹۷/۱)، png(۱/۸)، gif(۰/۹)، bmp(۰,۰۲)
درصد تصاویر قرار گرفته شده در برگ‌های ساختار سلسله مراتبی	۹۸/۶
درصد تصاویری که برای قرار گرفتن آن‌ها در ساختار سلسله مراتبی نیاز بوده است تا گره های جدیدی اضافه شود.	۲۱/۹

میانگین دقت ارزیابی شده برای هر کاربر در این فعالیت ۸۸/۵۳ % محاسبه شده است که میانگین دقت اندازه گیری شده برای تمامی کاربران شرکت‌کننده در جمع‌سپاری را نمایش می‌دهد. میانگین دقت ارزیابی شده برای هر کار ۸۹/۴۶% تخمین زده می‌شود. دقت میانگین برای هر کار به صورت فرمول(۱) محاسبه می‌شود.

$$AAJ_j = \frac{\sum_i UAT_{ij}}{Tasks} \qquad (۱)$$

که در آن AAJ_j دقت میانگین برای کار j و UAT_{ij} دقت میانگین کاربران برای وظیفه i از کار j و Tasks تعداد کل وظایف در کار j ام است. در مورد خطای کاربران در برچسب زنی این خطا عمدتاً انسانی است و ممکن است به دلایلی مانند عدم توجه کاربر انسانی در لحظه و یا عدم مطالعه‌ی توضیحات ارائه شده صورت گیرد. در موارد بسیار کمی مانند بناها و باغ های ایرانی نیز تصاویری که مربوط به نماهای داخلی ساختمان‌ها و یا کاشی کاری‌ها می باشند، نیاز به دقت زیاد و یا جستجو در جویشگرهایی مانند گوگل است تا کاربر بتواند نتایج درست را انتخاب کند. همچنین در مواردی مانند خوشنویسی ایرانی و اسلامی کاربر باید دقت بالایی را صرف تشخیص و انجام کار نماید. جدول ۷ خلاصه‌ای از اطلاعات مربوط به فرآیند جمع‌سپاری را نمایش می‌دهد.

جدول۷ خلاصه ای از اطلاعات جمع سپاری

شرح	مقدار
تعداد افراد شرکت کننده در مرحله‌ی نهایی جمع‌سپاری	۶۴ نفر
مدت زمان جمع‌سپاری	۲ هفته
تعداد تصاویر جمع‌سپاری شده	۲۲۸۵۲۶
تعداد تصاویر مرتبط حاصل از جمع‌سپاری	۷۱۸۷۳
میانگین دقت ارزیابی شده برای هر کاربر	% ۸۸/۵۳
میانگین دقت ارزیابی شده برای هر کار	% ۸۹/۴۶
تعداد ترادف‌های جمع سپاری شده	۱۰۰۰
میانگین تعداد تصاویر حاصل از جمع سپاری برای هر ترادف	۷۱

۵ آمار کلی و جزییاتی از تصویرنت

شبکه‌ی تصویری تصویرنت حدود ۸ میلیون تصویر را در قالب بیش از ۳۳ هزار ترادف در اختیار محققان بینایی ماشین و پردازش تصویر قرار می‌دهد. جدول ۸ جزئیاتی از این شبکه‌ی تصویری را نمایش می دهد.

جدول۸ آمار و تصاویر ترادف‌ها

شرح	مقدار
تعداد کل تصاویربارگیری شده	۷۸۹۰۷۴۵
تعداد کل تصاویر حاصل از جمع‌سپاری	۷۱۸۷۳
تعداد ترادف‌های دارای تصویر مربوط به ImageNet	۲۱۲۹۵
تعداد ترادف‌های جدید ایرانی واسلامی	۱۰۰۴
تعداد کل تصاویر تصویرنت	۷۹۶۲۶۱۸
تعداد کل ترادف‌های تصویرنت	۳۳۲۹۹

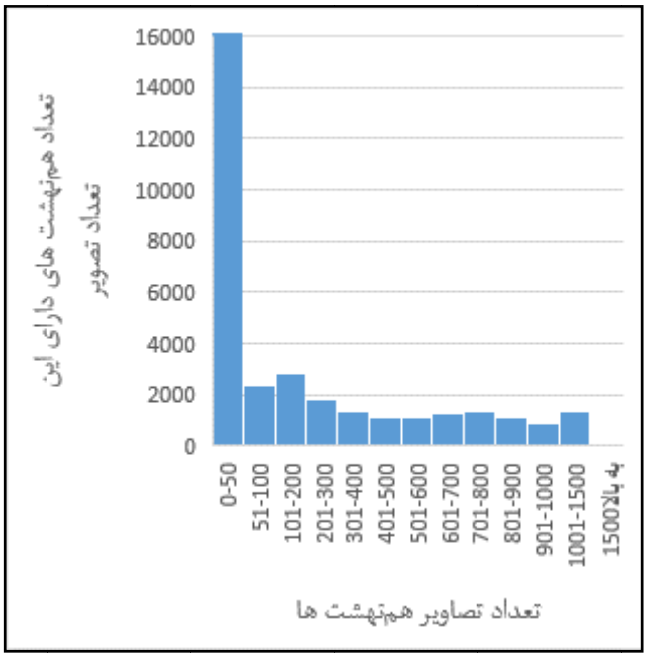
برای ۱۰۰۰ مفهوم در تصویرنت ۱۰۲۰ برچسب اضافه شده است. متوسط برچسب برای هر تصویر در تصویرنت ۱/۰۲ برچسب می‌باشد. این مقدار برای ImageNet ، ۱/۸ برچسب برای هر تصویر است. لازم به ذکر است که ImageNet شامل مجموعه‌ای از بسیاری از ترادف‌های رایج است که دارای معادل نیز هستند. ولی در مورد کلمات نادرتر نظیر آنچه که مختص زبان فارسی است وجود معادل‌های متفاوت برای یک ترادف کمتر محتمل است.

توابع کتابخانه‌ای برای استفاده از دارایی‌های تصویرنت که شامل اطلاعات ترادف‌ها، تصاویر و مرجع تصاویر می‌باشد، تعریف شده و در دسترس کاربران قرار گرفته است. توابع کتابخانه‌ای تصویرنت شامل دو بخش توابع داخلی و توابع خارجی می‌باشد. توابع کتابخانه‌ای داخلی، توابعی هستند که به صورت داخلی توسط ماژول‌های تصویرنت فراخوانی شده و با یکپارچه‌سازی نتایج و خروجی‌های آن‌ها به فراخوانی توابع خارجی پاسخ داده می‌شود. توابع کتابخانه‌ای خارجی نیز توابع رابط برنامه‌ی کاربردی هستند که توسط کاربران خارجی فراخوانی

می‌شوند. به دلیل حساسیت و امنیت توابع کتابخانه‌ای خارجی، نیاز است قابلیت کنترل بیشتر بر روی فراخوانی این توابع توسط کاربران بیرونی داشت. در همین راستا الگوی طراحی نماینده برای پیاده‌سازی این توابع کتابخانه‌ای خارجی استفاده شده است. به این ترادف که توابع کتابخانه‌ای خارجی نیز از همان توابع کتابخانه‌ای داخلی استفاده کرده ولی کنترل بیشتری بر روی فراخوانی آن‌ها فرآهم ساخته است.

در راستای هدف یک شبکه‌ی تصویری که فراهم نمودن حداکثر تعداد تصاویر ممکن برای هر ترادف آن است، تصویرنت توانسته است به طور میانگین ۲۳۹ تصویر را برای هر هم نهشت فراهم نماید در حالی که میانگین تعداد تصاویر در هر ترادف ImageNet تعداد ۴۳۹ تصویر می‌باشد. اما حدود ۱۶,۰۰۰ ترادف دارای صفر تا ۵۰ تصویر می‌باشند که ۱۱,۰۰۰ از این ترادف‌ها در نسخه‌ی اولیه بدون تصویر هستند. شکل ۵ توزیع تصاویر در ترادف‌های تصویرنت را نمایش می‌دهد.

تصویرنت با ارائه‌ی توابع کتابخانه ای امکان دسترسی و استفاده‌ی محققان پردازش تصویر از این شبکه‌ی تصویری را فراهم نموده است.



شکل ۵ نمودار توزیع تصاویر درترادف‌های تصویرنت

۶ کاربردهای شبکه تصویری

یک شبکه یادگیری عمیق بدون داشتن مقدار عظیم و ساختار یافته از تصاویر برای اسامی مختلف (شبکه تصویری) نمی‌تواند سامانه مناسبی برای تشخیص اشیا با قدرت تشخیص بالا ایجاد نماید. برای درک بهتر از ضرورت‌های شبکه‌ی تصویری مثال‌هایی از کاربردهای آن، در ادامه معرفی می‌شوند.

- رانندگی خودکار: یک عامل نرم افزاری برای آنکه بتواند به درستی رانندگی کند، علاوه بر درک از محیط اطراف خود و شناخت کامل نسبت به قوانین رانندگی، باید درک ویژه ای از اشیا پیرامونش داشته باشد. به عنوان مثال باید بتواند فرق کاغذی که مچاله شده و درخیابان افتاده را با یک قطعه سنگ به همان اندازه تشخیص دهد. این امر در تشخیص اینکه می تواند از روی آن رد شود یا نه کمک می کند. همچنین باید تفاوت انسان را با اشیا دیگر مانند درخت درک کند. در زمان تصادف بین برخورد با انسان و با یک درخت نیاز به تصمیم بلادرنگ و دقیق دارد.
- هواپیماهای بدون سرنشین: بدون در نظر گرفتن کاربرد هواپیما های بدون سرنشین (مانند کاربردهای نظامی، رهگیری تغییرات جنگل های بارانی، ساخت و ساز شهری و...) این هواپیماها باید بتوانند بین اشیاء مختلف که بر فراز آنها پرواز می کنند دید درستی داشته باشد.
- دوربین های امنیتی: یک دوربین امنیتی باید بتواند اشیاء را به درستی تشخیص دهد. سپس رفتار آن‌ها را بررسی و نیت و اعمال آنها را به درستی درک کند. به عنوان مثال یک دوربین امنیتی در یک استخر باید فرق یک کودک که در آب شنا می کند با دیگری که در حال غرق شدن است را تشخیص دهد. یک دوربین امنیتی باید یک سارق را در حال سرقت کردن تشخیص دهد.
- دوربین های ترافیکی و کنترل شهر: این دوربین‌ها باید دید وسیع و سطح بالایی از اتفاق‌ها و واقعیت های موجود داشته باشند. باید بتوانند رفتار های اجتماعی را در سطح شهر مشاهده و درک کنند. به عنوان مثال میزان خودروهای فرسوده در شهر را شناسایی کنند.
- ربات های امدادگر: در زمان جنگ، بلایای طبیعی، آتش سوزی و سایر حوادث زیان بار جانی به ربات های امداد گر نیاز است. ربات های امداد گر باید در تشخیص اشیا، توصیف شرایط و شدت وخامت مجروحان، یافتن اقدام مناسب در جهت بهبود مصدوم و انتخاب بهترین رویکرد در کمک به مجروحان همواره به بهترین شکل ممکن عمل کنند.
- کمک به پزشکان و پرستاران: در نگهداری بیماران، شناسایی بیماری و کنترل وضعیت بیمار، همواره چشم های بیدار و دقیق بهترین کمک را می کند. پرستاران نیاز به کنترل مداوم مریض ندارند. تنها در زمان نیاز به بیمار رسیدگی می کنند. پزشکان در تشخیص بیماری‌ها و درمان بیمار از ماشین‌ها استفاده میبرند.
- کمک به نابینایان: نابینایان همواره در انجام امور عادی خود در زندگی دچار مشکلات عدیده شده‌اند. طراحی ماشینی که بتواند ببیند و محیط را برای نابینایان توصیف کند یکی از با ارزش ترین خدمت های انسانی در این حوزه است.
- مدیریت و سازماندهی مفاهیم چند رسانه ای بر اساس ترادف: ImageNet ایده‌ی سازماندهی فایل های چندرسانه ای را به صورت سلسله مراتبی و بر اساس ترادف مطرح نموده است که مدیریت آن‌ها را نیز ساده می سازد.
- جویشگرها: تشخیص اشیا و استنباط منطقی و توصیف حالات از تصویر در زمان پرسش، تشخیص اشیا و استنباط منطقی و

توصیف حالات از تصویر در زمان خزش و نمایه گذاری، کاربرد در گسترش پرس و جوی تصویری و نگاشت ترادف به پرسش در جستجوی تصویر از جمله کاربردهای شبکه‌ی تصویری در جویشگرها می باشند.

۷ جمع‌بندی و آینده‌ی تصویرنت

تصویرنت نخستین شبکه‌ی تصویری ایرانی است که برای استفاده در جویشگرهای بومی و همچنین استفاده‌ی آسان در تحقیقات پردازش تصویر و بینایی ماشین تهیه شده است. تصویرنت با هدف بومی سازی شبکه‌ی تصویری ImageNet ایجاد گردید و تصاویر استفاده شده از ImageNet در آن از لینک های ارائه شده در ImageNet بدون نیاز به دادن تعهد به بوجودآورندگان ImageNet فراهم شده است تا بتوان از این مجموعه تصاویر برای کاربردهای تجاری نیز استفاده نمود. حال آنکه برای پایگاه داده تصاویری که بوسیله درخواست از ImageNet فراهم می گردد، چنین امکانی وجود ندارد و باید تنها برای کاربردهای غیر تجاری و آموزشی مورد استفاده قرار گیرند. به این ترتیب در این شبکه تصویری ۷۸۹۰۷۴۵ تصویر برای ۳۲۲۹۵ ترادف فارسی شده از **ImageNet** تهیه گردیده است.علاوه بر این بیش از ۱۰۰۰ ترادف مخصوص فرهنگ ایرانی و اسلامی به همراه ۷۱۸۷۳ تصویر مرتبط برای آن‌ها که با روش جمع‌سپاری برچسب‌گذاری شده اند، به تصویرنت اضافه شده است.

همانطور که مطرح شد یکی از جنبه های استفاده از شبکه‌ی تصویری کاربرد آن در جویشگرهای تصویری به منظور درک و تشخیص بهتر ترادف‌های مورد پرس‌وجو می باشد. از جهت دیگر یکی از اهداف تولید جویشگر بومی توسعه‌ی محتوای بومی و اشاعه‌ی فرهنگ ایرانی اسلامی می باشد. با توجه به اینکه شبکه‌ی تصویری به عنوان یک زیرساخت داده ای مدنظر قرار می گیرد وImageNet بر اساس شبکه‌ی کلمات انگلیسی وردنت ساخته شده است، برای کاربرد در جویشگر بومی نیاز به استفاده از خط و زبان فارسی و داده های مرتبط با فرهنگ ایرانی اسلامی حس می شود. در واقع علاوه بر مستقل شدن از استفاده‌ی مستقیم از ImageNet نیاز به پیاده سازی این بستر داده ای برای فرهنگ ایرانی اسلامی و زبان فارسی هستیم. در فرهنگ ایرانی اسلامی بسیاری از ترادف‌ها و اشیا وجود دارند که به هیچ‌وجه در پروژه اصلی ImageNet به کار برده نشده است. به عنوان مثال ترادف‌هایی مانند کرسی (لحاف کرسی)،سجاده، هفت سین، سمنو و ده‌ها مورد دیگر هستند که نیاز به گنجانده شدن در این شبکه تصویری برای آن‌ها وجود دارد و به تصویرنت اضافه شده اند.

شاید این سوال مطرح شود که چه نیازی به گنجاندن ترادف‌ها و واژگان فارسی است. اگرچه برخی عامل‌های نرم‌افزاری مثل دوربین‌هایی که سارق را تشخیص می دهند نیازی به بومی‌سازی ندارند. چون سارق در همه فرهنگ‌ها یک ترادف است، اما مثال‌های دیگری هستند که نیاز به بومی‌سازی دارند. به عنوان مثال به مورد کمک به نابینایان رجوع می کنیم. یک نابینا در ایران دنیا را طوری می‌خواهد ببیند که یک ایرانی می‌بیند. باید عامل نرم افزاری بتواند نماز خواندن را درک کند و آن را برای نابینا توصیف کند. باید بتواند مراسم از آتش پریدن چهارشنبه سوری را با واقعه آتش‌سوزی تفکیک کند. باید بتواند سفره هفت سین را نسبت به سفره شام تشخیص دهدو هزاران مورد دیگر که مختص فرهنگ ایرانی و اسلامی می‌باشد.

کارهای آتی در جهت تکامل تصویرنت شامل موارد زیر است: افزایش تعداد تصاویر این شبکه‌ی تصویری علی‌الخصوص تصاویر مربوط به ترادف‌های ایرانی و اسلامی، پالایش تصاویر تصویرنت برای تطابق هرچه بیشتر با شئونات ایرانی‌واسلامی، افزودن امکاناتی مانند تصاویر آنکادر شده که در آن‌ها اشیاء اصلی مربوط به تصویر مشخص شده است.

مراجع

[۱] رحمانی، فرزانه، حافظی، علی، زرگری، فرزاد، "**مروری بر ضرورت‌ها و کاربردهای شبکه‌ی تصویری**"، اولین کنفرانس ملی جویشگر بومی، ۳۷-۴۰، پژوهشگاه ارتباطات و فناوری اطلاعات(مرکز تحقیقات مخابرات ایران)، تهران، ۱۳۹۴.

[2] Miller, G. A., Backwith, R., Fellbaum C., Gross, D., Miller, K., *"Introduction to WordNet: An on-line lexical database"*, International Journal of Lexicography, Vol. 3, No. 4, pp. 235-244, 1990.

[3] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L., "*ImageNet: A large-scale hierarchical Image Database*", CVPR, 2009.

[4] Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J., Ng, A. Y., *" Building high-level features using large scale unsupervised learning"*, in Proceedings of the 29 th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.

[5] Dean, J., Corrado, G. S., Monga, R., Chen, K., Devin, M. Le, Q. V., Mao, M. Z., Ranzato, M. A., Senior, A. Senior, Tucker, P., Yang, K., Ng, A. Y., *"Large scale distributed deep networks"*, NIPS, 2012.

[۶] زرگری، فرزاد، **ایجاد شبکه تصویری برای فارسنت**، ۷۰۵۷/ص/۵۰۰/پ ، طرح جویشگر بومی، پژوهشگاه ارتباطات و فناوری اطلاعات، ۱۳۹۴-۱۳۹۵.

[۷] پرکار، زینب، رحمانی، فرزانه، زابل زاده، محمدحسین، زرگری، فرزاد، قاسمی، احسان، **گزارش نهایی پروژه ایجاد شبکه تصویری برای فارسنت (تصویرنت)**، پژوهشگاه ارتباطات و فناوری اطلاعات(مرکز تحقیقات مخابرات ایران)، پژوهشکده فناوری اطلاعات، گروه سکوهای فناوری اطلاعات، ۱۳۹۵.

[8] http://targoman.com/

[9] https://translate.google.com/

[10] https://www.wikipedia.org/
[11] Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Famian, A., Bagherbeigi, S., Fekri, E., Monshizadeh, M., Assi, S. M., "*Semi-Automatic Development of Farsnet; the Persian Wordnet*". Proceedings of 5th Global WordNet Conference (GWA2010). Mumbai, India, 2010.
[۱۲] مرکز دائرةالمعارف بزرگ اسلامی، دانشنامه ایران، ، جلد ۱-۴، مرکز دائره المعارف بزرگ اسلامی، ۱۳۸۴.
[۱۳] ، مرکز دائرةالمعارف بزرگ اسلامی، دائرةالمعارف بزرگ اسلامی، جلد ۱-۲۲، ۱۳۶۷ تا ۱۳۹۴.
[14] https://images.google.com/
[15] https://www.bing.com/images/
[16] http://image.parsijoo.ir/

فرزاد زرگری مدرک کارشناسی خود را در رشته مهندسی برق از دانشگاه صنعتی شریف، کارشناسی ارشد و دکتری تخصصی خود را از دانشگاه تهران دریافت کرده است. او اینک دانشیار دپارتمان فناوری اطلاعات در مرکز تحقیقات ICT، موسوم به مرکز تحقیقات مخابرات ایران می‌باشد. همچنین ایشان در دانشگاه آزاد اسلامی واحد علوم و تحقیقات مشغول به تدریس می‌باشد. زمینه‌های تحقیقاتی وی شامل سیستم‌های مالتی‌مدیا، پردازش‌تصویر و ویدیو و پیاده‌سازی سخت‌افزاری استانداردهای کدینگ تصویر و ویدیو است.

فرزانه رحمانی دانش‌آموخته کارشناسی مهندسی کامپیوتر نرم‌افزار دانشگاه شهید باهنر کرمان و کارشناسی ارشد مهندسی کامپیوتر دانشگاه تربیت مدرس می‌باشد. وی در حال حاضر دانشجوی دوره دکتری در پژوهشگاه ارتباطات و فناوری اطلاعات است و زمینه‌های تحقیقاتی مورد علاقه ایشان پردازش‌تصویر و ویدئو، استانداردهای فشرده سازی ویدئو و بازیابی تصویر و ویدئو است.

مژگان فرهودی دانش‌آموخته رشته مهندسی فناوری اطلاعات از دانشگاه صنعتی امیرکبیر می‌باشد. وی از سال ۱۳۸۰ در زمینه‌های بازیابی اطلاعات با تمرکز بر جویشگرها و سامانه های پرسش و پاسخ، وب‌کاوی و پردازش هوشمند زبان طبیعی فعالیت دارد. علاوه بر فعالیت‌های پژوهشی، ایشان در حال حاضر بعنوان مدیر گروه سکوهای فناوری اطلاعات و مجری آزمایشگاه ارزیابی خدمات وب (وب‌آزما) در پژوهشگاه ارتباطات و فناوری اطلاعات در حال فعالیت می‌باشند.

محمدحسین زابل‌زاده دانش‌آموخته کارشناسی مهندسی کامپیوتر نرم‌افزار دانشگاه آزاد اسلامی می‌باشد. وی در حال حاضر در زمینه رایانش ابری، و ارائه خدمات اینترنتی بر بستر ابر فعالیت دارد. مهمترین فعالیت حرفه‌ای ایشان در زمینه راه‌اندازی، طراحی و توسعه زیرساخت رایانش ابری و همچنین توسعه نرم‌افزارهای آماده برای اجرا بر زیرساخت ابر می‌باشد. از فعالیت‌های فرعی ایشان نیز می‌توان به طراحی گرافیک و همچنین طراحی و توسعه نرم‌افزارهای تحت وب اشاره کرد.

زینب پرکار دوره کارشناسی را در سال ۱۳۹۳ در دانشگاه آزاد اسلامی واحد پرند در رشته‌ی مهندسی کامپیوتر- نرم‌افزار به پایان رسانید و در همان سال در گرایش فناوری اطلاعات برای اخذ درجه‌ی کارشناسی ارشد در دانشگاه آزاد اسلامی واحد علوم تحقیقات تهران پذیرفته شد. علاقه‌مندی‌های پژوهشی وی، پردازش و نهان نگاری تصویر و موتورهای جستجوی تصویری هستند.

احسان قاسمی دانش آموخته کارشناسی مهندسی کامپیوتر نرم‌افزار دانشگاه آزاد اسلامی تهران شمال است. وی در حال حاضر دانشجوی دوره کارشناسی ارشد در دانشگاه علوم و تحقیقات است و زمینه‌های تحقیقاتی مورد علاقه ایشان رایانش‌ابری، امنیت و اپلیکیشن‌های وب و وب‌سوکت‌ها و نیز یادگیری عمیق است.