Arch

JCWR2020

کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۲ خرداد ۱۳۹۹

A Tourism Recommender System Based on Location and Semantic Web

Asghar Hoshyar[1], Mostafa Sabzekar [2]

[1] Master of Computer Engineering, Islamic Azad University, Birjand Branch, Iran

[2] Assistant Professor, Computer Engineering Department, Birjand University of Technology, Birjand, Iran
sabzekar@birjandut.ac.ir

Abstract

Tourism is one of the most important industries in the world. It plays an important role in the cultural and economic prosperity of countries. Therefore, the design and production of tourism recommendation systems has become one of the most popular topics for researchers in order to satisfy tourists. In this paper, a tourism recommendation system is proposed. In this system, tourists' interests are extracted by semantics using word clustering and emotion analysis. Finally, recommendations based on the interests of tourists are presented to them. The evaluation results of the proposed system provide high values in the evaluation criteria.

Keywords: Tourism Recommender System, Semantic Web, Text Analysis, Emotion Analysis

www.SID.ir

Arch

JCWR2020
کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۱ خرداد ۱۳۹۹

ارائه سیستم توصیه گر گردشگری مبتنی بر مکان با استفاده از وب معنایی

[۱] اصغر هشیار و [۲] مصطفی سبزه‌کار

[۱] دانشجوی کارشناسی ارشد تجارت الکترونیک، دانشگاه آزاد بیرجند

[۲] استادیار گروه مهندسی کامپیوتر، دانشگاه صنعتی بیرجند
sabzekar@birjandut.ac.ir

چکیده

گردشگری یکی از صنایع مهم جهان به شمار می‌رود و نقش به‌سزایی را در شکوفایی فرهنگی و اقتصادی کشورها ایفا می‌کند. از این رو، طراحی و تولید سیستم‌های توصیه‌گر گردشگری به یکی از موضوعات محبوب محققان به منظور رفاه حال گردشگران و جلب رضایت آن‌ها مبدل شده است. در این مقاله نیز یک سیستم توصیه‌گر گردشگری مورد بررسی قرار گرفته است. در این سیستم علایق گردشگران با استفاده از خوشه‌بندی لغات و تحلیل احساسات با بهره گیری از روش‌های معنایی استخراج می‌شوند. در نهایت توصیه‌هایی بر اساس علایق گردشگران به آن‌ها ارائه می‌گردد. نتایج ارزیابی سیستم پیشنهادی، نشان دهنده مقادیر بالایی در معیارهای ارزیابی است.

کلمات کلیدی

سیستم توصیه گر گردشگری، وب معنایی، تحلیل متن، تحلیل احساسات.

۱- مقدمه

امروزه گردشگری به عنوان صنعتی مهم و سودآور مورد توجه بسیاری از کشورها و مدیران جوامع قرار گرفته است. همزمان با افزایش اهمیت این صنعت، حجم بسیار زیادی از اطلاعات درباره جاذبه‌های گردشگری مختلف توسط گردشگران در فضای مجازی به وجود آمده است. این امر موجب پیچیدگی فرآیند تصمیم‌گیری افراد به منظور تعیین مقاصد گردشگری شده است [۱].

بنا به دلایل فوق، سیستم‌های توصیه‌گر گردشگری متفاوتی جهت سهولت تصمیم‌گیری گردشگران توسط محققان ایجاد شده‌است [۲]. اولین دسته از این سیستم‌ها، سیستم‌های مبتنی بر عکس‌هایی با برچسب جغرافیایی[۱] هستند. سیستم‌های موجود در این دسته با استخراج الگوی حرکتی گردشگران بر اساس موقعیت جغرافیایی آن‌ها، توصیه‌هایی را برای ادامه مسیر ارائه می‌دهند [۳]. به عنوان مثال در پژوهشی محققان بر اساس عکس‌های منتشر شده توسط افراد، گردشگرانی را که مقاصد گردشگری یکسانی دارند، در گروه‌های مشابه طبقه‌بندی شدند و توصیه‌هایی بر اساس سابقه بازدید سایر افراد هم گروهی، ارائه می‌دهند [۴]. در این روش‌ها به دلیل عدم توجه به نظرات کاربران و احساسات آن‌ها نتایج از دقت خوبی برخوردار نیستند.

به دلیل اهمیت نظرات افراد درباره جاذبه‌های گردشگری بازدید شده، دسته دیگری از سیستم‌های توصیه‌گر بر اساس نظرات گردشگران به وجود آمدند. به عنوان نمونه در یکی از ابتدایی‌ترین از این نوع تحقیقات محققان یک صفحه چت خصوصی مابین نماینده آژانس مسافرتی و کاربران ایجاد کردند و در آن سوالاتی از کاربران پرسیدند. آن‌ها پس از طراحی یک هستی‌شناسی منحصر به فرد گردشگری، در جواب کاربران به دنبال لغات موجود در هستی شناسی رفتند و بر این اساس سعی در تشخیص ترجیحات کاربران داشتند تا توصیه‌هایی مشابه ترجیحات را ایجاد کنند [۵].

در هر یک از دسته‌های فوق پژوهشگران متعددی سعی بر ارائه سیستم-های توصیه‌گر نموده‌اند. به طور مثال، در [۶] محققان با تاکید بر اینکه سیستم‌های توصیه‌گر امروزه نه تنها از منظر پژوهش، بلکه از دید اقتصادی قابل تامل هستند، به ارائه روشی برای توصیه یک سیستم گردشگری پرداخته است. نویسندگان در این مقاله توصیه یک مکان به گردشگر را با استفاده از یافتن مشابهت‌ها در سلیقه آنها، اعتماد و ارتباط آنها با یکدیگر در شبکه‌های

www.SID.ir

اجتماعی، انجام می‌دهند. در [۷] نیز یک سیستم توصیه‌گر مبتنی بر مکان گرشگری با لحاظ کردن مواردی مانند زمان، فصل، رفتار گردشگر و تاریخچه گرشگری‌های یک فرد ارائه شده است. در پژوهشی دیگر [۸] محققان سعی بر ارائه یک سیستم توصیه‌گر شخصی مبتنی بر مکان آگاه به زمینه ارائه شده است.

روش پیشنهادی این مقاله در دسته دوم (سیستم‌های مبتنی بر نظرات کاربر) قرار دارد. در این روش نظرات کاربر در قالب یک مجموعه‌ی متنی جمع‌آوری می‌شود و پس از گذراندن فرآیند پیش پردازش اسامی موجود در آن ها استخراج می‌شود. در مرحله بعد خوشه بندی معنایی بر روی نظرات انجام شده و ترجیحات کاربر با استفاده تحلیل احساسات استخراج می‌شوند. در نهایت با استفاده از ترجیحات کاربر و ویژگی‌های جاذبه‌های گردشگری، جاذبه‌هایی که بیشترین میزان شباهت با ترجیحات کاربر دارند پیشنهاد می‌شوند.

در ادامه ساختار مقاله به شرح زیر است: در بخش دوم روش پیشنهادی معرفی شده است و در بخش سوم نیز نتایج ارزیابی این روش مورد بررسی قرار گرفته است.

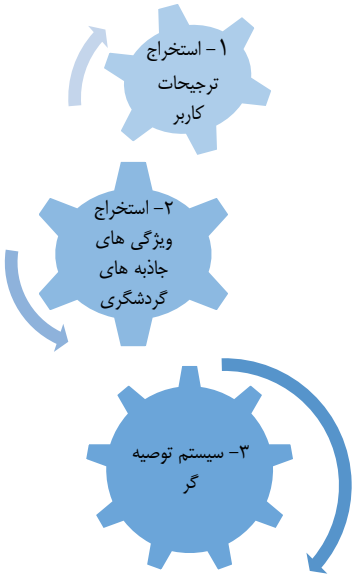
۲- روش پیشنهادی

روش پیشنهادی این مقاله دارای ۳ مرحله کلی می‌باشد که در شکل (۱) آمده است.

در ادامه هر یک از این مراحل به تفصیل مورد بررسی قرار می‌گیرند.

۲-۱- استخراج ترجیحات کاربر

فرآیند اصلی این مرحله در شکل (۲) آمده است. ورودی این مرحله مجموعه نظرات کاربر و خروجی آن ترجیحات کاربر است.



شکل (۱) : مراحل کلی روش پیشنهادی

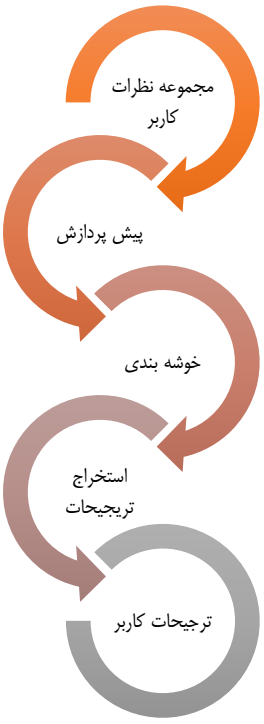
۲-۱-۱- مرحله اول: پیش پردازش

در این مرحله نظرات ثبت شده توسط کاربر به بخش‌های معناداری تقسیم می‌شوند. ابتدا دنباله‌ی کاراکترها توسط فرآیند واحد سازی به دنباله‌ای از کلمات تبدیل می‌شوند. در ادامه با استفاده از برچسب گذاری اجزای کلام[۲] نقش هر یک از کلمات (اعم از اسم، فعل، قید و ..) در جملات تعیین می‌شوند. پس از تعیین نقش کلمات، لغات توقف (لغات پرتکراری که دارای اهمیت و اطلاعات خاصی نیستند. مانند: a، the، and) حذف می‌شوند.

به دلیل کاهش حجم لغات ذخیره شده، لغات باقی مانده از مراحل قبل به ریشه اصلی خود برمی‌گردند [۹]. به عنوان مثال لغات eaten و eating به ریشه اصلی خود یعنی لغت eat بازگشته و ذخیره می‌شوند. در نهایت از میان لغات موجود، لغاتی که برچسب «اسم» دارند، ذخیره می‌شوند. این کار باعث بهبود کارایی خوشه بندی می‌شود [۱۰].

۲-۱-۲- مرحله دوم: خوشه بندی

در مرحله دوم با استفاده از روش *k-menas* به خوشه بندی لغات مرحله اول پرداخته می‌شود. در اولین گام از این مرحله ماتریسی از اسامی تشکیل می‌شود که سطرها و ستون‌های آن را اسامی مرحله اول تشکیل می‌دهند و درایه‌های آن را میزان شباهت معنایی مابین اسامی تشکیل می‌دهد. معیار شباهت معنایی مورد استفاده در این مقاله معیار محبوب وو-پالمر[۳] است [۱۱].



شکل (۲) : مراحل کلی استخراج ترجیحات در روش پیشنهادی

www.SID.ir

به دلیل اهمیت میزان وقوع کلمات در کنار یکدیگر، علاوه بر ماتریس معنایی ماتریس هم وقوعی لغات نیز تشکیل می‌شود [۱۲]. سطرها و ستون-های این ماتریس را لغات موجود در مجموعه نظرات تشکیل می‌دهند. درایه-های این ماتریس نیز بر اساس تعداد ظهور هر لغت با سایر لغات موجود در نظرات تکمیل می‌شوند.

در ادامه به ماتریس اسامی وزن α و β به ماتریس هم وقوعی لغات وزن β داده می‌شود (این اوزان بر اساس سعی و خطا و میزان اهمیت هر ماتریس در سناریوهای متفاوت تعیین می‌شوند). مجموع اوزان α و β برابر با یک است. در نهایت مقادیر ماتریس شباهت و ماتریس هم وقوعی لغات جمع شده و ماتریس نهایی تشکیل می‌شود. این ماتریس که در شکل (۳)) نیز آمده است، ماتریسی متقارن است.

در گام بعدی با استفاده از ماتریس نهایی، گرافی ایجاد می‌شود. که راس-های آن را اسامی و یال‌های آن در صورت وجود شباهت معنایی مابین لغات ایجاد می‌شوند. وزن هر یال برابر با میزان شباهت بین دو راس(اسم) است. در ادامه یال‌هایی که وزنی کمتر از حد آستانه دارند، حذف می‌شوند. پس از حذف یال‌های بی اهمیت گراف اولیه به تعدادی زیرگراف(خوشه) تقسیم می‌شود.

۲-۱-۳- مرحله سوم: استخراج ترجیحات

در این مرحله تمامی جملات بر اساس لغاتی که ازآن تشکیل شده‌اند، در خوشه‌های مربوطه قرار می‌گیرند. به عنوان مثال اگر کلمات یک جمله در دو خوشه قرار داشته باشند، این جمله در هر دو خوشه قرار می‌گیرد. در ادامه وارد فرآیند تحلیل احساسات می‌شویم. به شناسایی حالت‌های ذهنی افراد اعم از احساسات و عقاید و اعتقادات آن‌ها، تحلیل احساسات گفته می‌شود [۱۳]. در روش تحلیل احساسات این مقاله، پس از طی مراحل پیش‌پردازش و فرآیند برچسب گذاری گفتار، لغاتی با برچسب صفت، فعل و قید استخراج می‌شوند. در ادامه امتیازهایی برای هر یک از سه بار احساسی مترادف‌های لغات مدنظر تعیین می‌گردند [۱۴]. امتیازات احساسی مثبت، منفی و خنثی برای لغت w مشابه فرمول (۱) تعیین می‌شود:

$$\begin{cases} score_{pos}(w) = \frac{1}{n}\sum_{i=1}^{n} score_{pos} \\ score_{neg}(w) = \frac{1}{n}\sum_{i=1}^{n} score_{neg} \\ score_{obj}(w) = \frac{1}{n}\sum_{i=1}^{n} score_{obj} \end{cases} \quad (۱)$$

در این روش امتیاز مثبت از تقسیم مجموع بار مثبت مترادف‌ها(سینست-ها)ی لغت بر تعداد مترادف‌ها بدست می‌آید. امتیازات خنثی و منفی نیز بدین صورت تعیین می‌گردند. در ادامه و پس از تعیین سه امتیاز برای لغات، مجموع آن‌ها برای هر جمله محاسبه می‌شود. در صورتی که مجموع بار مثبت بیشتر از بار منفی باشد، بار کلی جمله منفی خواهد بود، در غیر این صورت جمله با احساسی منفی تلقی می‌شود. همچنین در صورتی که جنبه خنثی بودن جمله بیشترین امتیاز را کسب کند، این جمله به دلیل اهمیت پایین در نظر گرفته

	Noun 1	Noun 2	⋯	Noun n
Noun 1		X		W
Noun 2	X			Z
⋮				
Noun n	W	Z		

شکل (۳) : ماتریس متقارن نهایی

نخواهد شد. در نهایت مجموع امتیاز هر خوشه که شامل تعدادی از جملات است، از طریق فرمول (۲) محاسبه می‌شود:

$$S_c = \frac{N_p - N_n}{N_p + N_n} \quad (۲)$$

در فرمول فوق، تعداد جملات مثبت موجود در هر خوشه با عبارت N_p و تعداد جملات منفی هر خوشه با عبارت N_n نشان داده شده است و در نهایت پس از تعیین امتیازات هر خوشه، لغات خوشه‌ای که بیشترین امتیاز را دریافت کرده باشد به عنوان ترجیحات کاربر تعیین می‌شوند.

۲-۲- استخراج جاذبه های ویژگی‌های گردشگری

در این بخش به منظور تعیین و استخراج جاذبه‌های گردشگری، با استفاده از موقعیت مکانی فعلی کاربر، جاذبه‌هایی که در نزدیکی کاربر قرار گرفته‌اند، تعیین می‌شوند. در ادامه مجموعه نظراتی که توسط تمامی گردشگران ثبت شده است، جمع‌آوری می‌شوند. پس از طی مراحل فرآیند پیش‌پردازش بر روی این مجموعه، تعداد پنج اسمی که بیشترین تکرار را در مجموعه نظرات داشته باشند به عنوان ویژگی‌های جاذبه مدنظر تعیین می‌شوند.

۲-۳- سیستم توصیه‌گر

در این مرحله که مهم‌ترین بخش از سیستم توصیه‌گر پیشنهادی است، توصیه‌هایی به تناسب ترجیحات کاربر ارائه می‌شوند. ورودی این بخش ترجیحات کاربر و ویژگی جاذبه‌های گردشگری می‌باشند. در ادامه میزان شباهت لغات موجود در خوشه ترجیحات با لغات پرتکرار مستخرج از جاذبه-های گردشگری تعیین می‌شودو در نهایت جاذبه‌های گردشگری اطراف کاربر بر اساس بیشترین میزان مشابهت به ترتیب به کاربر جهت بازدید توصیه می-شوند. معیار محاسبه شباهت معنایی مورد استفاده در این رساله معیار وو-پالمر می‌باشد که در بخش اول به تفصیل بیان شده است.

۳- ارزیابی و نتیجه‌گیری

پیاده سازی روش پیشنهادی با استفاده از زبان پایتون انجام شده است. به منظور ارزیابی داده‌های حاوی نظرات گردشگران از سایت TripAdvisor [۱۵] استخراج شده است. مجموعه داده‌ی آموزشی شامل نظرات پنجاه

www.SID.ir

گردشگر با سنین و ملیت‌های متفاوت درباره جاذبه‌های گردشگری مختلف در مدت ۶ ماه(ماه‌های آوریل تا دسامبر سال ۲۰۱۸) است. مجموعه داده تست نیز شامل نظرات گردشگر در اولین سفر بعد دسامبر ۲۰۱۸ است. حد آستانه مورد استفاده در مرحله خوشه‌بندی نیز با استفاده از سعی و خطا به مقدار ۷۵ درصد انتخاب شده است.

در ارزیابی روش پیشنهادی، برای جاذبه‌های بازدید شده پس از دسامبر ۲۰۱۸ پنج کلمه پرتکرار از میان نظرات گردشگران درباره هر جاذبه استخراج شده است. سپس سفر به جاذبه‌هایی که کلمات پرتکرار آن‌ها شباهتی بیش از ۸۰ درصد با ترجیحات گردشگران داشته است، به عنوان موفقیت روش پیشنهادی در نظر گرفته شده است.

پس از مقایسه میزان دقت معیارهای متفاوت شباهت معنایی اعم از کسینوسی، جاکارد و وو–پالمر مشخص شد که معیار وو–پالمر دارای بیشترین دقت در خوشه بندی در روش پیشنهادی است. با در نظر گرفتن شعاع های متفاوت(شعاع a به اندازه مساحت هر شهر، شعاع b به اندازه نصف مساحت هر شهر و شعاع c به اندازه یک چهارم مساحت هر شهر) جاذبه‌های گردشگری در مسافت‌های متفاوت بررسی می‌شوند. مقادیر precision برای سیستم پیشنهادی در صورتی که خروجی سیستم یک پیشنهاد، دو پیشنهاد و سه پیشنهاد باشد، در نمودار (۱) آمده است.

با توجه به نتایج، precision به میزانی که محدوده انتخابی اطراف کاربر بیشتر می‌شود، افزایش می‌یابد. پس از بررسی، نتایج نشان می‌دهد که میزان precision همزمان با افزایش تعداد توصیه‌های ارائه شده و شعاع مطلوب کاربر افزایش یافته است. افزایش احتمال مراجعه به جاذبه‌های پیشنهادی همزمان با افزایش تعداد توصیه‌ها دلیل این اتفاق می‌باشد.

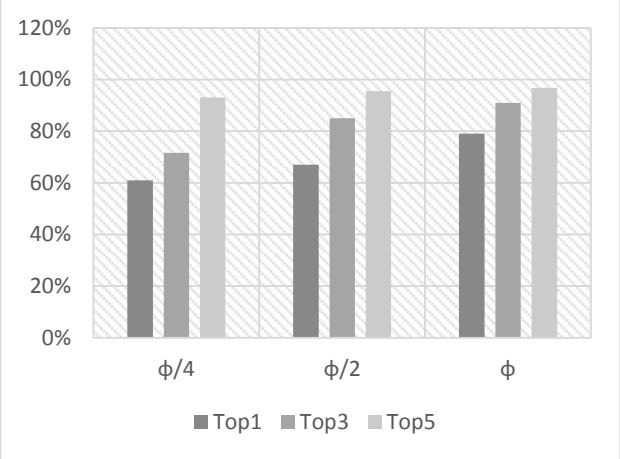
مقادیر recall نیز برای سیستم پیشنهادی در صورتی که خروجی سیستم یک پیشنهاد، دو پیشنهاد و سه پیشنهاد باشد در نمودار (۲) آمده است.

نمودار (۲) نشان دهنده کاهش مقادیر recall در صورت افزایش شعاع مطلوب گردشگر است. کاهش جاذبه‌های بازدید شده در برابر تعداد توصیه‌های ارائه شده توسط سیستم پیشنهادی دلیل این اتفاق است.

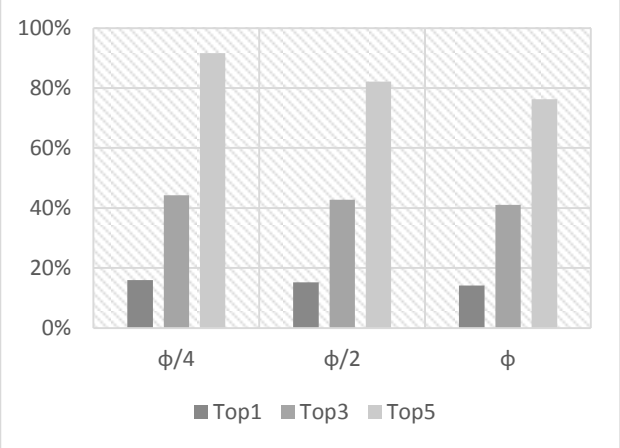
در نهایت مقادیر f-measure نیز برای سیستم پیشنهادی در تعداد پیشنهادهای متفاوت در نمودار (۳) آمده است.

در نهایت افزایش مقادیر f-measure همزمان با افزایش تعداد پیشنهادات، نتایجی است که با توجه به نمودار (۳) بدست آمده است. عملکرد بهتر سیستم پیشنهادی در شعاع‌های کوتاه‌تر در صورت ارائه توصیه‌های بیشتر، نتیجه دیگری است که از این نمودار حاصل شده است.

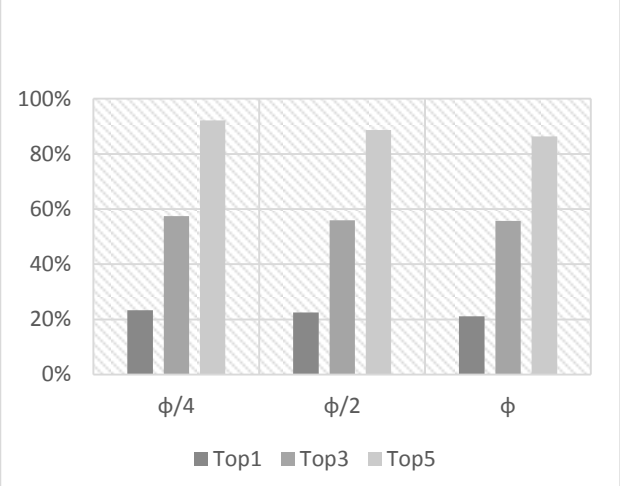
نتایج ارزیابی نشان دهنده موفقیت این روش در کسب مقادیر بالا در معیارهای ارزیابی(precision , recall و f-measure) است. در ادامه پیشنهاد می‌شود با مجهز کردن سیستم پیشنهادی به اطلاعات زمینه‌ای اعم از وضعیت آب و هوایی و تعیین ترافیک در مسیر بازدید از جاذبه‌ها، کیفیت توصیه‌ها را افزایش داد. یکی دیگر از راهکارهای بهبود سیستم پیشنهادی مجهز کردن به زبان‌های دیگری اعم از فارسی می‌باشد. در این صورت



نمودار (۱) : مقدار precision در مسافت‌ها و تعداد پیشنهادهای متفاوت



نمودار (۲) : مقدار recall در مسافت‌ها و تعداد پیشنهادهای متفاوت



نمودار (۳) : مقدار f-measure در مسافت‌ها و تعداد پیشنهادهای متفاوت

www.SID.ir

Arch

سیستم پیشنهادی می‌تواند به صورت بومی نیز در تمامی کشورها مورد استفاده قرار بگیرد.

مراجع

[1] T. Hayashi and T. Yoshida, "Development of a Tour Recommendation," in *International Conference on Management Science and Engineering Management*, 2019. Kanazawa, Japan.

[2] R. Anacleto, L. Figueiredo, A. Almeida, and P. Novais, "Mobile application to provide personalized sightseeing tours," *Journal of Network and Computer Applications*, vol. 41, no. 1, 2014.

[3] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, , and G. Chen, "Travel recommendation using geo-tagged photos in social media for tourist." *Wireless Personal Communications*, vol. *80, no.* 4, 2015.

[4] A. Majid, L. Chen, G. Chen, H. T. Mirza, I. Hussain, and J. Woodward, "A context-aware personalized travel recommendation system based on geotagged social media data mining," *International Journal of Geographical Information Science.*, vol. 27, no. 4, 2013.

[5] S. Loh, F. Lorenzi, R. Saldaña, and D. Licthnow, "a Tourism Recommender System Based on Collaboration and Text Analysis," *Information Technology & Tourism*, vol. 6, no. 3, 2004.

[6] L. Esmaeili, S. Mardani, A. Hashemi Golpayegani, Z. Zanganeh Madar, "A novel tourism recommender system in the context of social commerce," Expert Systems with Applications, vol. 149, pp. 113301, 2020.

[7] T. Chatzidimitris, D. Gavalas, V. Kasapakis, *et al.* "A location history-aware recommender system for smart retail environments," Pers Ubiquit Comput, 2020, https://doi.org/10.1007/s00779-020-01374-7.

[8] R. Logesh and V. Subramaniyaswamy, "Exploring Hybrid Recommender Systems for Personalized Travel Applications," In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore, 2018.

[9] A. Garcia-Pablos, M. Cuadros, and M. T. Linaza, "OpeNER: open tools to perform natural language processing on accommodation reviews," in *Information and Communication Technologies in Tourism 2015*, Switzerland: Springer, 2015,

[10] S. Fodeh, B. Punch, and P. N. Tan, "On ontology-driven document clustering using core semantic features," *Knowledge and information systems*, vol. 28, no. 2, 2011.

[11] Z. Wu and M. Palmer, "Verbs semantics and lexical selection," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, 1994, pp. 133–138. Las Cruces, New Mexico.

[12] H. Geng, P. Zhao, E. Chen, Q. Cai, "A Novel Automatic Text Summarization Study Based term Co-Occurrence", In the Proceeding of the 5th International conference on Cognitive Informatics , 2006, Beijing, China.

[13] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, 2016.

[14] Denecke, Kerstin. "Using sentiwordnet for multilingual sentiment analysis." *24th international conference on data engineering workshop*. IEEE, 2008.

[15] "TripAdvisor." [Online]. Available: www.maps.tripadvisor.com

زیرنویس‌ها

[1]Geo-tagged photos
[2]Part Of Speech(POS) Tagging
[3]Wu-Palmer

www.SID.ir